SIAM J. COMPUT.
Vol. 48, No. 2, pp. 581–643

CONVERGENCE OF MCMC AND LOOPY BP IN THE TREE UNIQUENESS REGION FOR THE HARD-CORE MODEL*

CHARILAOS EFTHYMIOU†, THOMAS P. HAYES‡, DANIEL ŠTEFANKOVIČ§, ERIC VIGODA¶, AND YITONG YIN‖

Abstract. We study the hard-core (gas) model defined on independent sets of an input graph where the independent sets are weighted by a parameter (aka fugacity) $\lambda > 0$. For constant Δ, the previous work of Weitz [*Proceedings of STOC*, 2006, pp. 140–149] established an FPTAS for the partition function for graphs of maximum degree Δ when $\lambda < \lambda_c(\Delta)$. Sly [*Proceedings of FOCS*, 2010, pp. 287–296] showed that there is no FPRAS, unless NP=RP, when $\lambda > \lambda_c(\Delta)$. The threshold $\lambda_c(\Delta)$ is the critical point for the statistical physics phase transition for uniqueness/nonuniqueness on the infinite Δ-regular tree. The running time of Weitz's algorithm is exponential in $\log \Delta$. Here we present an FPRAS for the partition function whose running time is $O^*(n^2)$. We analyze the simple single-site Markov chain known as the Glauber dynamics for sampling from the associated Gibbs distribution. We prove there exists a constant Δ_0 such that for all graphs with maximum degree $\Delta \geq \Delta_0$ and girth ≥ 7 (i.e., no cycles of length ≤ 6), the mixing time of the Glauber dynamics is $O(n \log n)$ when $\lambda < \lambda_c(\Delta)$. Our work complements that of Weitz, which applies for small constant Δ, whereas our work applies for all Δ at least a sufficiently large constant Δ_0. (This includes Δ depending on $n = |V|$.) Our proof utilizes loopy belief propagation (BP) which is a widely used algorithm for inference in graphical models. A novel aspect of our work is using the principal eigenvector for the BP operator to design a distance function which contracts in expectation for pairs of states that behave like the BP fixed point. We also prove that the Glauber dynamics behaves locally like loopy BP. As a byproduct we obtain that the Glauber dynamics, after a short burn-in period, converges close to the BP fixed point, and this implies that the fixed point of loopy BP is a close approximation to the Gibbs distribution. Using these connections we establish that loopy BP quickly converges to the Gibbs distribution when the girth ≥ 6 and $\lambda < \lambda_c(\Delta)$.

Key words. Gibbs sampling, hard-core model, Markov chain Monte Carlo, rapid mixing, loopy belief propagation

AMS subject classifications. Primary, 68W20; Secondary, 60K35, 82B20

DOI. 10.1137/17M1127144

1. Introduction.

1.1. Background. The hard-core gas model is a natural combinatorial problem that has played an important role in the design of new approximate counting algorithms and for understanding computational connections to statistical physics

*Received by the editors April 24, 2017; accepted for publication (in revised form) February 13, 2018; published electronically April 30, 2019. This work was done in part while the authors were visiting KAIST (Korea Advanced Institute of Science and Technology) and the Simons Institute for the Theory of Computing.
http://www.siam.org/journals/sicomp/48-2/M112714.html
Funding: The first author was supported by DFG grant EF 103/1-1. The third author was supported in part by NSF grant CCF-1318374. The fourth author was supported in part by NSF grants CCF-1217458 and CCF-1555579. The fifth author was supported by the National Science Foundation of China under grants 61672275 and 61722207.

†Goethe University, Frankfurt am Main, Germany (efthymiou@gmail.com).
‡Department of Computer Science, University of New Mexico, Albuquerque, NM 87131 (hayes@cs.unm.edu).
§Department of Computer Science, University of Rochester, Rochester, NY 14627 (stefanko@cs.rochester.edu).
¶School of Computer Science, Georgia Institute of Technology, Atlanta GA 30332 (ericvigoda@gmail.com).
‖State Key Laboratory for Novel Software Technology, Nanjing University, China (yinyt@nju.edu.cn).

phase transitions. For a graph $G = (V, E)$ and a fugacity $\lambda > 0$, the hard-core model is defined on the set Ω of independent sets of G, where $\sigma \in \Omega$ has weight $w(\sigma) = \lambda^{|\sigma|}$. The equilibrium state of the system is described by the Gibbs distribution μ in which an independent set σ has probability $\mu(\sigma) = w(\sigma)/Z$. The partition function $Z = \sum_{\sigma\in\Omega} w(\sigma)$.

We study the closely related problems of efficiently approximating the partition function and approximate sampling from the Gibbs distribution. These problems are important for Bayesian inference in graphical models where the Gibbs distribution corresponds to the posterior or likelihood distributions. Common approaches used in practice are Markov chain Monte Carlo (MCMC) algorithms, e.g., see [4, 25, 40], and message passing algorithms, such as loopy belief propagation (BP), e.g., see [27], and one of the aims of this paper is to prove fast convergence of these algorithms.

Exact computation of the partition function is #P-complete [38], even for restricted input classes [9], and hence the focus is on designing an efficient approximation scheme, either a deterministic FPTAS or randomized FPRAS. The existence of an FPRAS for the partition function is polynomial-time interreducible to approximate sampling from the Gibbs distribution.

A beautiful connection has been established: there is a computational phase transition on graphs of maximum degree Δ that coincides with the statistical physics phase transition on Δ-regular trees. The critical point for both of these phase transitions is $\lambda_c(\Delta) := (\Delta - 1)^{\Delta-1}/(\Delta - 2)^{\Delta}$. In statistical physics, $\lambda_c(\Delta)$ is the critical point for the uniqueness/nonuniqueness phase transition on the infinite Δ-regular tree $\mathbb{T}_\Delta$ [18]. (Roughly speaking, this is the phase transition for the decay versus persistence of the influence of the leaves on the root.) For some basic intuition about the value of this critical point, note its asymptotics $\lambda_c(\Delta) \sim e/(\Delta - 2)$ and the following basic property: $\lambda_c(\Delta) > 1$ for $\Delta \le 5$ and $\lambda_c(\Delta) < 1$ for $\Delta \ge 6$.

Weitz [42] showed, for all constant Δ, an FPTAS for the partition function for all graphs of maximum degree Δ when $\lambda < \lambda_c(\Delta)$. To properly contrast the performance of our algorithm with Weitz's algorithm let us state his result more precisely: for all $\delta > 0$, there exists constant $C = C(\delta)$, for all Δ, all $G = (V, E)$ with maximum degree Δ, all $\lambda < (1 - \delta)\lambda_c(\Delta)$, all $\epsilon > 0$, there is a deterministic algorithm to approximate Z within a factor $(1 \pm \epsilon)$ with running time $O((n/\epsilon)^{C \log \Delta})$. An important limitation of Weitz's result is the exponential dependence on $\log \Delta$ in the running time. Hence it is polynomial-time only for constant Δ, and even in this case the running time is unsatisfying.

On the other side, Sly [34] (extended in [6, 7, 35, 8]) has established that, unless $NP = RP$, for all $\Delta \ge 3$, there exists $\gamma > 0$, for all $\lambda > \lambda_c(\Delta)$, there is no polynomial-time algorithm for triangle-free Δ-regular graphs to approximate the partition function within a factor $2^{\gamma n}$.

Weitz's algorithm was extremely influential: many works have built upon his algorithmic approach to establish efficient algorithms for a variety of problems (e.g., [29, 32, 19, 20, 33, 39, 21, 31, 22]). One of its key conceptual contributions was showing how decay of correlations properties on a Δ-regular tree are connected to the existence of an efficient algorithm for graphs of maximum degree Δ. We believe our paper enhances this insight by connecting these same decay of correlations properties on a Δ-regular tree to the analysis of widely used MCMC and message passing algorithms.

1.2. Main results. As mentioned briefly earlier on, there are two widely used approaches for the associated approximate counting/sampling problems, namely, MCMC and message passing approaches. A popular MCMC algorithm is the simple single-site update Markov chain known as the Glauber dynamics. The Glauber dynamics is a Markov chain (X_t) on Ω whose transitions $X_t \rightarrow X_{t+1}$ are defined by the following process:

1. Choose v uniformly at random from V.
2. If $N(v) \cap X_t = \emptyset$, then let

$$X_{t+1} = \begin{cases} X_t \cup \{v\} & \text{with probability } \lambda/(1+\lambda), \\ X_t \setminus \{v\} & \text{with probability } 1/(1+\lambda). \end{cases}$$

3. If $N(v) \cap X_t \neq \emptyset$, then let $X_{t+1} = X_t$.

The mixing time T_{mix} is the number of steps to guarantee that the chain is within a specified (total) variation distance of the stationary distribution. In other words, for $\epsilon > 0$,

$$T_{\text{mix}}(\epsilon) = \min\{t : \forall X_0, d_{\mathsf{TV}}(X_t, \mu) \leq \epsilon\},$$

where $d_{\mathsf{TV}}()$ is the variation distance. We use $T_{\text{mix}} = T_{\text{mix}}(1/4)$ to refer to the mixing time for $\epsilon = 1/4$.

It is natural to conjecture that the Glauber dynamics has mixing time $O(n \log n)$ for all $\lambda < \lambda_c(\Delta)$. Indeed, Weitz's work implies rapid mixing for $\lambda < \lambda_c(\Delta)$ for amenable graphs. On the other hand Mossel, Weitz, and Wormald in [26] show slow mixing when $\lambda > \lambda_c(\Delta)$ on random regular bipartite graphs. The previously best known results for MCMC algorithms are far from reaching the critical point. It was known that the mixing time of the Glauber dynamics (and other simple, local Markov chains) is $O(n \log n)$ when $\lambda < 2/(\Delta - 2)$ for any graph with maximum degree Δ [5, 23, 40]. In addition, [13] analyzed Δ-regular graphs with $\Delta = \Omega(\log n)$ and presented a polynomial-time simulated annealing algorithm when $\lambda < \lambda_c(\Delta)$.

Here we prove $O(n \log n)$ mixing time up to the critical point when the maximum degree is at least a sufficiently large constant Δ_0, and there are no cycles of length ≤ 6 (i.e., girth ≥ 7).

THEOREM 1. *For all $\delta > 0$, there exists $\Delta_0 = \Delta_0(\delta)$ and $C = C(\delta)$, and for all graphs $G = (V, E)$ of maximum degree $\Delta \geq \Delta_0$ and girth ≥ 7, all $\lambda < (1-\delta)\lambda_c(\Delta)$, all $\epsilon > 0$, the mixing time of the Glauber dynamics satisfies*

$$T_{\text{mix}}(\epsilon) \leq Cn \log(n/\epsilon).$$

Note that Δ and λ can be a function of $n = |V|$. The above sampling result yields (via [36, 15]) an FPRAS for estimating the partition function Z with running time $O^*(n^2)$, where $O^*()$ hides multiplicative $\log n$ factors. The algorithm of Weitz [42] is polynomial-time for small constant Δ; in contrast our algorithm is polynomial-time for all $\Delta > \Delta_0$ for a sufficiently large constant Δ_0.

A family of graphs of particular interest are random Δ-regular graphs and random Δ-regular *bipartite* graphs. These graphs do not satisfy the girth requirements of Theorem 1 but they have few short cycles. Hence, as one would expect the above result extends to these graphs.

THEOREM 2. *For all $\delta > 0$, there exists $\Delta_0 = \Delta_0(\delta)$ and $C = C(\delta)$, and for all $\Delta \geq \Delta_0$, all $\lambda < (1-\delta)\lambda_c(\Delta)$, all $\epsilon > 0$ and with probability $1 - o(1)$ over the choice of an n-vertex graph G chosen uniformly at random from the set of all Δ-regular graphs, the mixing time of the Glauber dynamics on G satisfies*

$$T_{\text{mix}}(\epsilon) \leq Cn \log(n/\epsilon).$$

The same holds for G chosen uniformly at random from the set of all Δ-regular bipartite graphs.

Theorem 2 complements the work in [26], which shows slow mixing for random Δ-regular bipartite graphs when $\lambda > \lambda_c(\Delta)$.

Theorem 2 is essentially a corollary of Theorem 1 and its proof amounts to relaxing the girth restriction to having a limited number of short cycles in the neighborhood of a vertex.

To prove Theorem 1 we have to analyze the well-known BP algorithm. BP, introduced by Pearl [28], is a simple recursive scheme designed on trees to correctly compute the marginal distribution for each vertex to be occupied/unoccupied. In particular, consider a rooted tree $T = (V, E)$, where for $v \in V$ its parent is denoted as p and its children are $N(v)$. Let

$$q(v) = \mathbf{Pr}_\mu\left[v \text{ is occupied} \mid p \text{ is unoccupied}\right]$$

denote the probability in the Gibbs distribution that v is occupied conditional on its parent p being unoccupied. It is convenient to work with ratios of the marginals, and hence let $R_{v\to p(v)} = q(v)/(1 - q(v))$ denote the ratio of the occupied to unoccupied marginal probabilities. Because T is a tree it is not difficult to show that this ratio satisfies the following recurrence:

$$R_{v\to p(v)} = \lambda \prod_{w\in N(v)\setminus\{p(v)\}} \frac{1}{1 + R_{w\to v}}.$$

This recurrence explains the terminology of BP that $R_{w\to v}$ is a "message" from w to its parent v. Given the messages to v from all of its children then v can send its message to its parent. Finally the root r (with a parent p always fixed to be unoccupied and thus removed) can compute the marginal probability that it is occupied by $q(r) = R_{r\to p}/(1 + R_{r\to p})$.

The above formulation defines (the sum-product version of) BP a simple, natural algorithm which works efficiently and correctly for trees. For general graphs *loopy BP* implements the above approach, even though there are now cycles and so the algorithm no longer is guaranteed to work correctly. For a graph $G = (V, E)$, for $v \in V$ let $N(v)$ denote the set of all neighbors of v. For each $p \in N(v)$ and time $t \geq 0$ we define a message

$$R^t_{v\to p} = \lambda \prod_{w\in N(v)\setminus\{p\}} \frac{1}{1 + R^{t-1}_{w\to v}}.$$

The corresponding estimate of the marginal can be computed from the messages by

$$q^t(v, p) = \frac{R^t_{v\to p}}{1 + R^t_{v\to p}}. \tag{1}$$

Loopy BP is a popular algorithm for estimating marginal probabilities in general graphical models (e.g., see [27]), but there are few results on when loopy BP converges to the Gibbs distribution (e.g., Weiss [41] analysed graphs with one cycle, and [37, 14, 16] presented various sufficient conditions; see also [2, 30] for analysis of BP variants).

We show that the Glauber dynamics behaves locally like loopy BP. Furthermore, we show that loopy BP converges to a unique fixed point for $\lambda < \lambda_c(\Delta)$. Combining together these two facts allows us to characterize the local behavior of the Glauber dynamics in terms of the fixed point of loopy BP. This is a key element in our proof of Theorem 1.

The following result is a byproduct of the analysis of Theorem 1; we feel it is of independent interest. We show that loopy BP converges quickly to the Gibbs distribution. More specifically, we prove that, on any graph with girth ≥ 6 and maximum degree $\Delta \geq \Delta_0$, where Δ_0 is a sufficiently large constant, loopy BP quickly converges to the (marginals of) Gibbs distribution μ. More precisely, $O(1)$ iterations of loopy BP suffices; note each iteration of BP takes $O(n+m)$ time where $n = |V|$ and $m = |E|$.

THEOREM 3. *For all $\delta, \epsilon > 0$, there exists $\Delta_0 = \Delta_0(\delta, \epsilon)$ and $C = C(\delta, \epsilon)$, and for all graphs $G = (V, E)$ of maximum degree $\Delta \geq \Delta_0$ and girth ≥ 6, all $\lambda < (1-\delta)\lambda_c(\Delta)$, the following holds: for $t \geq C$, for all $v \in V$, $p \in N(v)$,*

$$\left| \frac{q^t(v,p)}{\mu(v \textit{ is occupied} \mid p \textit{ is unoccupied})} - 1 \right| \leq \epsilon,$$

where $\mu(\cdot)$ is the Gibbs distribution.

1.3. Contributions. Our main conceptual contribution is formally connecting the behavior of BP and the Glauber dynamics. We will analyze the Glauber dynamics using path coupling [1]. In path coupling we need to analyze a pair of *neighboring configurations*; in our setting this is a pair of independent sets X_t, Y_t which differ at exactly one vertex v with $X_t(v), Y_t(v)$ being unoccupied and occupied, respectively. The key is to construct a one-step coupling $(X_t, Y_t) \to (X_{t+1}, Y_{t+1})$ and introduce a distance function $\mathcal{D} : \Omega \times \Omega \to \mathbf{R}_{\geq 0}$ which "contracts," meaning that the following *path coupling condition* holds for some $\gamma > 0$:

$$\mathbb{E}\left[\mathcal{D}(X_{t+1}, Y_{t+1}) \mid X_t, Y_t\right] \leq (1-\gamma)\mathcal{D}(X_t, Y_t).$$

We use a distance function of the form $\mathcal{D}(X_t, Y_t) = \sum_{v \in X_t \oplus Y_t} \Phi(v)$ for an appropriate weighting $\Phi : V \to \mathbf{R}_{\geq 1}$. That is, the distance $\mathcal{D}(X_t, Y_t)$ is a sum over disagreeing vertices and each disagreement v contributes weight $\Phi(v)$.

We use a simple maximal one-step coupling and hence in our setting the path coupling condition simplifies to

$$(1-\gamma)\Phi(v) \geq \sum_{z \in N(v)} \frac{\lambda}{1+\lambda} \mathbf{1}\left\{z \text{ is unblocked in } X_t\right\} \Phi(z), \tag{2}$$

where *unblocked* means that $N(z) \cap X_t = \emptyset$, i.e., all neighbors of z are unoccupied, and we have assumed there are no triangles so as to ignore the possibility that X_t and Y_t differ on the neighborhood of z.

The distance function $\mathcal{D}$ must satisfy a few basic conditions such as being a path metric, and if $X \neq Y$, then $\mathcal{D}(X, Y) \geq 1$ (so that by Markov's inequality $\mathbf{Pr}\left[X_t \neq Y_t\right] \leq \mathbb{E}\left[\mathcal{D}(X_t, Y_t)\right]$). A standard choice for the distance function is the Hamming distance. In our setting the Hamming distance does not suffice and our primary challenge is determining a suitable distance function.

We cannot construct a suitable distance function which satisfies the path coupling condition for arbitrary neighboring pairs X_t, Y_t. To this end, we utilize the loopy BP recurrences corresponding to the probability that a vertex is unblocked. A key insight is that we can show the existence of a suitable Φ for the distance function $\mathcal{D}$ when the local neighborhood of the disagreement v behaves like the BP fixed point: roughly, in (2) the number of unblocked vertices in $N(v)$ is equal to what we expect to have if each neighbor is occupied with probability specified by the BP fixed point.

We feel our construction of this distance function $\mathcal{D}$ is our most interesting contribution. Note that the relevant qualitative information for the neighbors of v is whether or not they are unblocked (rather than simply unoccupied). Hence consider the (unrooted) BP recurrences corresponding to the probability that a vertex is unblocked. This corresponds to the following function $F : [0,1]^V \to [0,1]^V$, which is defined as follows, for any $\omega \in [0,1]^V$ and $z \in V$:

$$F(\omega)(z) = \prod_{y \in N(z)} \frac{1}{1+\lambda\omega(y)}. \tag{3}$$

Also, for some integer $i \geq 0$, let $F^i(\omega) : [0,1]^V \to [0,1]^V$ be the i-iterate of F. This recurrence is closely related to the standard BP operator $R()$ and hence under the hypotheses of our main results, we have that $F()$ has a unique fixed point ω^*, and for any ω, all $z \in V$, $\lim_{i\to\infty} F^i(z) = \omega^*(z)$.

To construct the distance function $\mathcal{D}$ we start with the Jacobian of this BP operator $F()$. Since $F()$ converges to a fixed point, and, in fact, it contracts at every level w.r.t. an appropriately defined potential function, we then know that the Jacobian of the BP operator $F()$ evaluated at its fixed point ω^* has spectral radius < 1. What motivates the use of BP is the observation that with a suitable similarity transformation of the Jacobian we obtain a matrix which encodes the following path coupling condition when the pair of states is close to the BP fixed points, namely,

$$\Phi(v) > \sum_{z \in N(v)} \frac{\lambda\omega^*(z)}{1+\lambda\omega^*(z)}\Phi(z). \tag{4}$$

This captures the main idea in the construction of a suitable Φ. However to apply path coupling additional requirements are needed for Φ. For example, to measure the rate of contraction we need to bound the gap of the principal eigenvalue from 1, and to apply Markov's inequality we need that $\mathcal{D}(X,Y) \geq 1$ when $X \neq Y$. Hence additional technical work is required to explicitly derive a Φ that behaves similar to the principal eigenvector. For a comparison between (2) and (4), it is useful to recall that Δ is assumed to be large, i.e., $\Delta > \Delta_0$, while for such Δ the fugacity λ behaves as $\lambda = O(1/\Delta)$.

There are previous works [11, 12] which utilize the spectral radius of the adjacency matrix of the input graph G to design a suitable distance function for path coupling. In contrast, we use insights from the analysis of the BP operator to derive a suitable distance function. We believe this is a richer connection that can potentially lead to stronger results since it directly relates to convergence properties on the tree. Our approach has the potential to apply for a more general class of spin systems; we comment on this in more detail in the conclusions.

The above argument only implies that we have contraction in the path coupling condition for pairs of configurations which are BP fixed points, i.e., the number of unblocked neighbors of the disagreeing vertex v is $\approx \sum_{z \in N(v)} \omega^*(z)$. A priori we don't even know if the BP fixed points on the tree correspond to the Gibbs distribution on the input graph. We prove that the Glauber dynamics (approximately) satisfies a recurrence that is close to the BP recurrence; this builds upon ideas of Hayes [10] for colorings. This argument requires that there are no cycles of length ≤ 6 for the Glauber dynamics (and no cycles of length ≤ 5 for the direct analysis of the Gibbs distribution). The girth requirements are used to prove (rough) independence of the probability that neighbors of a vertex v are unblocked, and hence concentration results can be utilized; this is explained in further detail in section 4.1. Some local sparsity condition is necessary since if there are many short cycles then the Gibbs distribution no longer behaves similarly to a tree and hence loopy BP may be a poor estimator.

As a consequence of the above relation between BP and the Glauber dynamics, we establish that from an arbitrary initial configuration X_0, after a short burn-in period of $T = O(n \log \Delta)$ steps of the Glauber dynamics the configuration X_T is a close approximation to the BP fixed point. In particular, for any vertex v, the number of unblocked neighbors of v in X_T is $\approx \sum_{z \in N(v)} \omega^*(z)$ with high probability. As is standard for concentration results, our proof of this result necessitates that Δ is at least a sufficiently large constant. Finally we adapt ideas of [4] to utilize these burn-in properties and establish rapid mixing of the Glauber dynamics. This essentially outlines the proof of Theorem 1.

Choose the initial configuration X_0 of the Glauber dynamics to be from the Gibbs distribution of the hard-core model. Then, X_T for $T = O(n \log \Delta)$ is not only related to the BP fixed points as we described in the previous paragraph, but it is also distributed as the Gibbs distribution. This observation makes it apparent that there should be a relation between the fixed point of loopy BP and the Gibbs distribution of the hard-core model. Theorem 3 provides a more systematic treatment of this observation. Essentially, it proves that for every vertex u, $\omega^*(u)$ is very close to the Gibbs marginal of u being unblocked. Theorem 3 instead of dealing with equilibrium configurations of the Glauber dynamics it considers samples from the hard-core model. That is, it establishes a relation between samples of the Gibbs distribution of the hard-core model and loopy BP which is similar to that of the Glauber dynamics. The use of samples from the Gibbs distribution instead of the Glauber dynamics allows us to improve the girth dependence to ≥ 6.

We prove Theorem 3 by means of the BP equations in (3), which are unrooted recurrences for being unblocked. We chose this specific system of equations because of its similarity to the recurrences that arise in the analysis for local uniformity, and its applicability for the path coupling analysis. Note that there are two key differences between the recurrences in (3) and (1). First, (3) captures unblocked whereas (1) captures unoccupied. The second difference is the significant one: the equations in (3) consider an unrooted version of BP while those in (1) consider a rooted version. There is a simple transformation between the two *rooted* versions: the rooted analog of the unblocked recurrences defined in (3) and the rooted, unoccupied recurrences defined in (1). In particular the corresponding fixed points for unblocked and unoccupied differ by a factor of λ. However, considering the unrooted version of unblocked is different than its rooted version. For $\lambda < \lambda_c(\Delta)$, the difference in the fixed points is $O(1/\Delta)$. Since our proof requires $\Delta > \Delta_0$ for a sufficiently large constant Δ_0 in order for concentration bounds, the error we introduce by considering the unrooted version is of a smaller order than the error ϵ we have in Theorem 3. Therefore, we can utilize the simpler system (namely, unrooted) with no additional loss in the quality of results we prove.

At this point it is worth pointing out why the lower bound $\Delta > \Delta_0$ is inevitable with our approach. We prove concentration results on the number of neighbors that are unblocked; this is an integer-valued function and hence we cannot obtain bounds closer than a factor $(1 \pm 1/\Delta)$. Therefore, as we require closer concentration bounds in order to get closer to the threshold we need that Δ grows.

1.4. Outline of paper. In the following section we state results about the convergence of the BP recurrences. Section 2 contains the proofs of all the results about the convergence of BP. We only postpone until section 7 the proofs of Propositions 7 and 8, which are a bit lengthy.

We present in section 3 our theorem showing the existence of a suitable distance function to use for path coupling; this is Theorem 9. The proof of Theorem 9 appears in section 3 and builds on the results of section 2.

Section 4 discusses local uniformity results for both the Glauber dynamics and the Gibbs distribution (these are Theorems 11 and 10); these uniformity results are necessary ingredients in the proofs of Theorems 1 and 3, respectively. The proof of Theorem 10 appears inside section 4, but the proof of Theorem 11 is more technical and lengthy and appears in sections 9 and 10.

Section 5 proves our main result, Theorem 1. For the proof we use the distance function we presented in section 3 and the uniformity result from section 4. In section 5.1 we provide a proof sketch of Theorem 1. In sections 5.3 and 5.4 we give some preliminary results and in section 5.2 we give the full proof of Theorem 1.

The extension of our rapid mixing result to random regular graphs (and regular, bipartite graphs) as stated in Theorem 2 is proved in section 6.

Theorem 3 about the efficiency of loopy BP is proved in section 8. (Some of the key technical results in the proof of Theorem 3 are already proved in sections 4 and 7.)

Sections 9 and 10 are the most technical parts of our paper; these sections prove the local uniformity property for the Glauber dynamics. In particular, section 9 provides some basic results and concepts we use in section 10 to prove uniformity.

Finally, section 11 provides some concluding remarks.

2. BP convergence. Here we state several useful results about the convergence of BP to a unique fixed point, and stepwise contraction of BP to the fixed point.

Our first lemma says that the recurrence for $F()$ defined in (3) has a unique fixed point.

LEMMA 4. *For all $\delta > 0$, there exists $\Delta_0 = \Delta_0(\delta)$, and for all $G = (V, E)$ of maximum degree $\Delta \geq \Delta_0$, all $\lambda < (1-\delta)\lambda_c(\Delta)$, the function F has a unique fixed point ω^*. Moreover, for any initial value $\omega^0 \in [0,1]^V$, denoting by $\omega^i = F^i(\omega)$ the vector after the ith iterate of F, it holds that*

$$\|\omega^i - \omega^*\|_\infty \leq 3(1-\delta/6)^i.$$

A critical result for our approach is that the recurrences $F()$ have stepwise contraction to the fixed point ω^*. To obtain contraction we use the following potential function Ψ. Let the function $\Psi : [0,1] \to \mathbb{R}_{\geq 0}$ be as follows:

$$\Psi(x) = \left(\sqrt{\lambda}\right)^{-1} \operatorname{arcsinh}\left(\sqrt{\lambda \cdot x}\right). \tag{5}$$

The following fact (which is formally verified in section 2.1) is frequently used in our analysis. For the λ and Δ assumed by Lemma 4, for any $x_1, x_2 \in [(1+\lambda)^{-\Delta}, 1]$,

$$\frac{1}{3}|x_1 - x_2| \leq |\Psi(x_1) - \Psi(x_2)| \leq 3|x_1 - x_2|. \tag{6}$$

Our main motivation for introducing Ψ is as a normalizing potential function that we use to define the following distance metric, D, on functions $\omega \in [0,1]^V$:

$$D(\omega_1, \omega_2) = \max_{z \in V} |\Psi(\omega_1(z)) - \Psi(\omega_2(z))|.$$

We will also need a variant, $D_{v,R}$, of this metric whose value only depends on the restriction of the function to a ball of radius ℓ around vertex v. For any $v \in V$, integer $\ell \geq 0$, let $B_\ell(v)$ be the set of vertices within distance $\leq \ell$ of v. Moreover, for functions $\omega_1, \omega_2 \in [0,1]^V$, we define

$$D_{v,\ell}(\omega_1, \omega_2) = \max_{z \in B(v,\ell)} |\Psi(\omega_1(z)) - \Psi(\omega_2(z))| . \tag{7}$$

We can now state the following convergence result for the recurrences, which establishes stepwise contraction.

Lemma 5. *For all $\delta > 0$, there exists $\Delta_0 = \Delta_0(\delta)$, and for all $G = (V,E)$ of maximum degree $\Delta \geq \Delta_0$, all $\lambda < (1-\delta)\lambda_c(\Delta)$, for any $\omega \in [0,1]^V$, $v \in V$, and $\ell \geq 1$, we have*

$$D_{v,\ell-1}(F(\omega), \omega^*) \leq (1-\delta/6) D_{v,\ell}(\omega, \omega^*),$$

where ω^* is the fixed point of F.

We also consider a recurrence which corresponds to the rooted belief propagation. For an undirected graph $G = (V,E)$, let $\bar{E}$ be the set of all orientations of edges in E. The function $H : [0,1]^{\bar{E}} \to [0,1]^{\bar{E}}$ is defined as follows: For any $\omega \in [0,1]^{\bar{E}}$ and $(v,p) \in \bar{E}$,

$$H(\omega)(v,p) = \prod_{u \in N(v) \setminus \{p\}} \frac{1}{1 + \lambda \omega(u,v)}. \tag{8}$$

For this system we have similar convergence result as Lemma 4.

Corollary 6. *For $G = (V,E)$ and λ assumed by Lemma* 4, *the function H defined in* (8) *has a unique fixed point ω^*. Moreover, for any initial value $\omega^0 \in [0,1]^{\bar{E}}$, denoting by $\omega^i = H^i(\omega)$ the vector after the ith iterate of H, it holds that*

$$\|\omega^i - \omega^*\|_\infty \leq 3(1-\delta/6)^i.$$

2.1. Proofs of Lemmas 4 and 5 and Corollary 6. We first verify the fact stated in (6). By the mean value theorem, for any $x_1, x_2 \in [(1+\lambda)^{-\Delta}, 1]$, there exists a mean value $\xi \in [(1+\lambda)^{-\Delta}, 1]$ such that

$$|\Psi(x_1) - \Psi(x_2)| = \Psi'(\xi)|x_1 - x_2| = \frac{1}{2\sqrt{\xi(1+\lambda\xi)}}|x_1 - x_2|.$$

Using a coarse estimation such that $\lambda_c(\Delta) < 3/(\Delta-2)$, it is easy to verify that for all sufficiently large Δ, we have $(1+\lambda)^{-\Delta} > (1+3/(\Delta-2))^{-\Delta} > 1/36$ and hence $\xi \in [1/36, 1]$, and also $\lambda < \lambda_c(\Delta) < 1/4$. Therefore, $\frac{1}{2\sqrt{\xi(1+\lambda\xi)}} \geq \frac{1}{2\sqrt{1+\lambda}} > \frac{1}{3}$ and $\frac{1}{2\sqrt{\xi(1+\lambda\xi)}} < \frac{1}{2\sqrt{\xi}} < 3$. This proves (6).

Next, we state two propositions which are needed for proving Lemmas 4 and 5. The proofs of these propositions use ideas from [29, 20, 32] and are postponed to section 7.

Let $f_{\lambda,d}(x) = (1+\lambda x)^{-d}$ be the symmetric version of the BP recurrence (3). Let $\hat{x} = \hat{x}(\lambda, d)$ be the unique fixed point of $f_{\lambda,d}(x)$, satisfying $\hat{x}(\lambda,d) = (1+\lambda\hat{x}(\lambda,d))^{-d}$. We define

$$\alpha(\lambda, d) = \sqrt{\frac{d \cdot \lambda\hat{x}(\lambda,d)}{1+\lambda\hat{x}(\lambda,d)}}. \tag{9}$$

PROPOSITION 7. *For all* $\delta > 0$, *there exists* $\Delta_0 = \Delta_0(\delta)$, *and for all* $\Delta \geq \Delta_0$, *all* $\lambda < (1-\delta)\lambda_c(\Delta)$, *where* $\lambda_c(\Delta) = \frac{(\Delta-1)^{\Delta-1}}{(\Delta-2)^{\Delta}}$, *it holds that* $\alpha(\lambda, \Delta) \leq 1 - \delta/6$.

Recall the function $F()$ as defined in (3). The following proposition was proved implicitly in [20].

PROPOSITION 8 (see [20]). *Let* $G = (V, E)$ *be a graph with maximum degree at most* Δ. *Assume that* $\alpha(\lambda, \Delta) \leq 1$. *For any* $\omega \in [0,1]^V$, *and* $v \in V$,

$$\sqrt{\frac{\lambda F(\omega)(v)}{1+\lambda F(\omega)(v)}} \sum_{u \in N(v)} \sqrt{\frac{\lambda \omega(u)}{1+\lambda \omega(u)}} \leq \alpha(\lambda, \Delta),$$

where $\alpha(\lambda, \Delta)$ *is defined in* (9).

By Proposition 7, for the regime of λ given in Lemmas 4 and 5, it holds that $\alpha(\lambda, \Delta) < 1 - \delta/6$, where Δ is the maximum degree of the graph $G = (V, E)$.

We then show that for any $\omega_1, \omega_2 \in [0,1]^V$ and $v \in V$,

$$|\Psi(\omega_1(v)) - \Psi(\omega_2(v))| \leq 1, \tag{10}$$

and

$$|\Psi(F(\omega_1)(v)) - \Psi(F(\omega_2)(v))| \leq (1-\delta/6) \max_{u \in N(v)} |\Psi(\omega_1(v)) - \Psi(\omega_2(v))|. \tag{11}$$

We first prove (10). It is easy to see that $\Psi(x)$ is monotonically increasing for $x \in [0,1]$, and thus $|\Psi(\omega_1(v)) - \Psi(\omega_2(v))| \leq \Psi(1) - \Psi(0) = \operatorname{arcsinh}(\sqrt{\lambda})/\sqrt{\lambda}$. Observe that $\operatorname{arcsinh}(x) \leq x$ for any $x \geq 0$ and hence $\operatorname{arcsinh}(\sqrt{\lambda})/\sqrt{\lambda} \leq 1$. Then (10) follows.

We then prove (11). Note that the derivative of the potential function Ψ is $\Psi'(x) = \frac{\mathrm{d}\,\Psi(x)}{\mathrm{d}\,x} = \frac{1}{2\sqrt{x(1+\lambda x)}}$. Due to the mean value theorem, there exists an $\tilde{\omega} \in [0,1]^{N(v)}$ such that

$$\begin{aligned}
&|\Psi(F(\omega_1)(v)) - \Psi(F(\omega_2)(v))| \\
&\quad = \sum_{u \in N(v)} \left| \frac{\partial F(\omega)(v)}{\partial \omega(u)} \bigg|_{\omega = \tilde{\omega}} \frac{\Psi'(F(\tilde{\omega})(v))}{\Psi'(\tilde{\omega}(u))} \right| |\Psi(\omega_1(u)) - \Psi(\omega_2(u))| \\
&\quad = \sqrt{\frac{\lambda F(\tilde{\omega})(v)}{1+\lambda F(\tilde{\omega})(v)}} \sum_{u \in N(v)} \sqrt{\frac{\lambda \tilde{\omega}(u)}{1+\lambda \tilde{\omega}(u)}} |\Psi(\omega_1(u)) - \Psi(\omega_2(u))| \\
&\quad \leq \left(\sqrt{\frac{\lambda F(\tilde{\omega})(v)}{1+\lambda F(\tilde{\omega})(v)}} \sum_{u \in N(v)} \sqrt{\frac{\lambda \tilde{\omega}(u)}{1+\lambda \tilde{\omega}(u)}} \right) \cdot \max_{u \in N(v)} |\Psi(\omega_1(u)) - \Psi(\omega_2(u))|.
\end{aligned}$$

Then (11) is implied by Propositions 7 and 8.

Now we are ready to prove Lemma 4. Consider the dynamical system defined by $\omega^{(i)} = F(\omega^{(i-1)})$ with arbitrary two initial values $\omega_1^{(0)}, \omega_2^{(0)} \in [0,1]^V$. The derivative of the potential function satisfies that $\Psi'(x) \geq \frac{1}{2\sqrt{1+\lambda}}$ for any $x \in [0,1]$. Due to the mean value theorem, for any $v \in V$, there exists a mean value $\xi \in [0,1]$ such that

$$\begin{aligned}
\left|\omega_1^{(i)}(v) - \omega_2^{(i)}(v)\right| &= \frac{1}{\Psi'(\xi)} \left| \Psi\left(\omega_1^{(i)}(v)\right) - \Psi\left(\omega_2^{(i)}(v)\right) \right| \\
&\leq 2\sqrt{1+\lambda} \left| \Psi\left(\omega_1^{(i)}(v)\right) - \Psi\left(\omega_2^{(i)}(v)\right) \right|.
\end{aligned}$$

Combined with (10) and (11), we have

$$\begin{aligned}\left\|\omega_1^{(i)}-\omega_2^{(i)}\right\|_\infty &\le 2\sqrt{1+\lambda}\left\|\Psi\left(\omega_1^{(i)}\right)-\Psi\left(\omega_2^{(i)}\right)\right\|_\infty\\ &\le 2(1-\delta/6)^i\sqrt{1+\lambda}\max_{z\in V}\left|\Psi\left(\omega_1^{(0)}(z)\right)-\Psi\left(\omega_2^{(0)}(z)\right)\right|\\ &\le 2(1-\delta/6)^i\sqrt{1+\lambda},\end{aligned}$$

which is at most $3(1-\delta/6)^i$ for $\lambda<\lambda_c(\Delta)$ for all sufficiently large Δ. Lemma 4 is proved.

Lemma 5 is then a consequence of this. According to the definition of $D_{v,R}$ in (7),

$$\begin{aligned}&D_{v,R-1}(F(\omega),\omega^*)\\ &\quad=\max_{u\in B(v,R-1)}|\Psi(F(\omega)(u))-\Psi(\omega^*(u))|\\ &\quad=\max_{u\in B(v,R-1)}|\Psi(F(\omega)(u))-\Psi(F(\omega^*)(u))| &&(\omega^*\text{ is fixed point})\\ &\quad\le\max_{u\in B(v,R-1)}(1-\delta/6)\max_{z\in N(u)}|\Psi(\omega(z))-\Psi(\omega^*(z))| &&(\text{due to (11)})\\ &\quad=(1-\delta/6)\max_{u\in B(v,R)}|\Psi(\omega(u))-\Psi(\omega^*(u))|\\ &\quad=(1-\delta/6)\cdot D_{v,R}(\omega,\omega^*),\end{aligned}$$

which proves Lemma 5.

Finally, with the approach used above, analyzing the convergence of H which is defined in (8) is the same as analyzing F on a graph G with maximum degree $\Delta-1$. Recall that $\alpha(\lambda,\Delta)$ is increasing in Δ. The same proof gives us Corollary 6.

3. Path coupling distance function. We now prove that there exists a suitable distance function Φ for which the path coupling condition holds for configurations that correspond to the fixed points of $F()$.

THEOREM 9. *For all $\delta>0$, there exists $\Delta_0=\Delta_0(\delta)$, and for all $G=(V,E)$ of maximum degree $\Delta\ge\Delta_0$, all $\lambda<(1-\delta)\lambda_c(\Delta)$, there exists $\Phi: V\to\mathbb{R}_{\ge 0}$ such that for every $v\in V$,*

$$1\le\Phi(v)\le 12 \tag{12}$$

and

$$(1-\delta/6)\Phi(v)\ge\sum_{u\in N(v)}\frac{\lambda\omega^*(u)}{1+\lambda\omega^*(u)}\Phi(u), \tag{13}$$

where ω^* is the fixed point of F defined in (3)*.*

The theorem is proved by considering the Jacobian J of the BP operator F:

$$J(v,u)\;=\;\left|\frac{\partial F(\omega)(v)}{\partial\omega(u)}\right|\;=\;\begin{cases}\frac{\lambda F(\omega)(v)}{1+\lambda\omega(u)} & \text{if } u\in N_v,\\ 0 & \text{otherwise.}\end{cases}$$

Let J^* denote the Jacobian at the fixed point $\omega=\omega^*$, formally:

$$J^*=J|_{\omega=\omega^*}. \tag{14}$$

Let D be the diagonal matrix with $D(v,v) = \omega^*(v)$ and define

$$\hat{J} = D^{-1} J^* D. \tag{15}$$

The path coupling condition (13) is in fact

$$\hat{J}\Phi \le (1-\delta/6)\Phi. \tag{16}$$

The fact that ω^* is a Jacobian attractive fixed point implies the existence of a nonnegative Φ with $\hat{J}\Phi < \Phi$. Thus, the theorem would follow immediately if the spectral radius of $\hat{J}$ is $\rho(\hat{J}) \le 1-\delta/6$ and $\hat{J}$ has a principal eigenvector with each entry from the bounded range $[1, 12]$. However, explicitly calculating this principal eigenvector can be challenging on general graphs.

The convergence of BP which is established in Lemmas 4 and 5, w.r.t. the potential function Ψ, guides us to an explicit construction of Φ to satisfy $\hat{J}\Phi < \Phi$ as follows.

Proof of Theorem 9. Due to Propositions 7 and 8, for any $\omega \in [0,1]^V$, and $v \in V$, we have

$$\sqrt{\frac{\lambda F(\omega)(v)}{1+\lambda F(\omega)(v)}} \sum_{u \in N(v)} \sqrt{\frac{\lambda\omega(u)}{1+\lambda\omega(u)}} \le 1-\delta/6.$$

In particular, this inequality holds for the fixed point ω^*, where $F(\omega^*)(v) = \omega^*(v)$. Therefore,

$$\sqrt{\frac{\lambda\omega^*(v)}{1+\lambda\omega^*(v)}} \sum_{u \in N(v)} \sqrt{\frac{\lambda\omega^*(u)}{1+\lambda\omega^*(u)}} \le 1-\delta/6.$$

Note that the derivative of the potential function Ψ is given by $\Psi'(x) = \frac{1}{2\sqrt{x(1+\lambda x)}}$. Therefore, the above inequality in fact gives us

$$\sum_{u \in N(v)} J^*(v,u) \frac{\Psi'(\omega^*(v))}{\Psi'(\omega^*(u))} = \sqrt{\frac{\lambda\omega^*(v)}{1+\lambda\omega^*(v)}} \sum_{u \in N(v)} \sqrt{\frac{\lambda\omega^*(u)}{1+\lambda\omega^*(u)}} \le 1-\delta/6,$$

which is equivalent to the following:

$$\sum_{u \in N(v)} \frac{\hat{J}(v,u)}{\omega^*(u)\Psi'(\omega^*(u))} \le \frac{1-\delta/6}{\omega^*(v)\Psi'(\omega^*(v))},$$

where in above J^* and $\hat{J}$ are as defined in (14) and (15).

Then, (16) is trivially satisfied by choosing Φ such that $\Phi(v) = \frac{1}{2\omega^*(v)\Psi'(\omega^*(v))} = \sqrt{\frac{1+\lambda\omega^*(v)}{\omega^*(v)}}$. In turn we get the path coupling condition (13).

Next, we show that this Φ satisfies that $1 \le \Phi(v) \le 12$. Since $\omega^* \in [0,1]^V$, we have $\Phi(v) = \sqrt{\frac{1+\lambda\omega^*(v)}{\omega^*(v)}} \ge 1$. Meanwhile, it holds that $\omega^*(v) = \prod_{u \in N_v} \frac{1}{1+\lambda\omega^*(u)} \ge (1+\lambda)^{-\Delta}$. By our assumption, $\lambda \le (1-\delta)\lambda_c(\Delta) \le \frac{4}{\Delta-2}$ for all $\Delta \ge 3$. Therefore, $\omega^*(v) \ge (1+\frac{4}{\Delta-2})^{-\Delta} \ge 5^{-3}$ and $\Phi(v) = \sqrt{\frac{1+\lambda\omega^*(v)}{\omega^*(v)}} \le \sqrt{5^3+4} \le 12$. ∎

4. Local uniformity. We will prove that the Glauber dynamics, after a sufficient burn-in, behaves with high probability locally similar to the BP fixed points. In this section we will formally state some of these "local uniformity" results.

For an independent set σ, for $v \in V$, and $p \in N(v)$ let

$$\mathbf{U}_{v,p}(\sigma) = \mathbf{1}\left\{\sigma \cap (N(v) \setminus \{p\}) = \emptyset\right\} \tag{17}$$

be the indicator of whether the children of v leave v unblocked.

We now state our main local uniformity results. We first establish that the Gibbs distribution behaves as in the BP fixed point, when the girth ≥ 6. We will prove that for any vertex v, the number of unblocked neighbors of v is $\approx \sum_{z\in N(v)} \omega^*(z)$ with high probability. Hence, for $v \in V$ let

$$\mathbf{S}_X(v) = \sum_{z\in N(v)} \mathbf{U}_{z,v}(X),$$

denote the number of unblocked neighbors of v in configuration X.

THEOREM 10 (Gibbs distribution uniformity). *For all $\delta, \epsilon > 0$, there exists $\Delta_0 = \Delta_0(\delta, \epsilon)$ and $C = C(\delta, \epsilon)$, and for all graphs $G = (V, E)$ of maximum degree $\Delta \geq \Delta_0$ and girth ≥ 6, all $\lambda < (1-\delta)\lambda_c(\Delta)$, for all $v \in V$, it holds that*

$$\mathbf{Pr}_{X\sim\mu}\left[\left|\mathbf{S}_X(v) - \sum_{z\in N(v)} \omega^*(z)\right| \leq \epsilon\Delta\right] \geq 1 - \exp\left(-\Delta/C_{10}\right),$$

where ω^* is the fixed point from Lemma 4.

Theorem 10 will be the key ingredient in the proof of Theorem 3. Before proving it let us give a brief discussion about its analogue for Glauber dynamics.

For our rapid mixing result (Theorem 2) we need an analogous local uniformity result for the Glauber dynamics. This will require the slightly higher girth requirement ≥ 7 since the grandchildren of a vertex v no longer have a certain conditional independence and we need the additional girth requirement to derive an approximate version of the conditional independence. (This is discussed in more detail in section 9.2.)

The path coupling proof weights the vertices according to Φ. Hence, in place of $\mathbf{S}$ we need the following weighted version $\mathbf{W}$. For $v \in V$ and $\Phi : V \to \mathbb{R}_{\geq 0}$ as defined in Theorem 9 let

$$\mathbf{W}_\sigma(v) = \sum_{z\in N(v)} \mathbf{U}_{z,v}(\sigma)\ \Phi(z). \tag{18}$$

We then prove that the Glauber dynamics, after sufficient burn-in, also behaves as in the BP fixed point with a slightly higher girth requirement ≥ 7. (For path coupling we only need an upper bound on the number of unblocked neighbors, and hence we state and prove this simpler form.)

THEOREM 11 (Glauber dynamics uniformity). *For all $\delta, \epsilon > 0$, let $\Delta_0 = \Delta_0(\delta, \epsilon)$, $C = C(\delta, \epsilon)$, and for all graphs $G = (V, E)$ of maximum degree $\Delta \geq \Delta_0$ and girth ≥ 7, all $\lambda < (1-\delta)\lambda_c(\Delta)$, let (X_t) be the Glauber dynamics on the hard-core model. For all $v \in V$, it holds that*

$$\mathbf{Pr}\left[(\forall t \in \mathcal{I}) \quad \mathbf{W}_{X_t}(v) < \sum_{z\in N(v)} \omega^*(z)\Phi(z) + \epsilon\Delta\right] \geq 1 - \exp\left(-\Delta/C\right), \tag{19}$$

where the time interval $\mathcal{I} = [Cn\log\Delta, n\exp(\Delta/C)]$.

The above theorem shows that for arbitrary initial state X_0 after $O(n \log \Delta)$ steps it achieves the local uniformity property whp (with high probability). Our proof of Theorem 1 will, in fact, use a slight variant of the above theorem, but the relevant notions haven't been presented yet so we defer the formal statement to Theorem 18 in section 5.3. Theorem 18 considers X_0 which is "nice" in an appropriate sense and proves that then only $O(n)$ steps are required to achieve the local uniformity property whp. Theorem 11 will then follow as a corollary of Theorem 18 (together with Lemma 21, which also appears in section 5.3).

4.1. Proof of Theorem 10. Here we present the proof of Theorem 10 of the local uniformity results for the Gibbs distribution.

Consider a graph $G = (V, E)$. For a vertex v and an independent set σ, consider the following quantity:

$$\mathbf{R}(\sigma, v) = \prod_{z \in N(v)} \left(1 - \frac{\lambda}{1+\lambda} \mathbf{U}_{z,v}(\sigma)\right), \tag{20}$$

where $\mathbf{U}_{z.v}(\sigma)$ is defined in (17). (It is the indicator that the children of z leave it unblocked.) The important aspect of this quantity $\mathbf{R}$ is the following interpretation. Let Y be distributed as in the Gibbs measure w.r.t. G. For triangle-free G we have

$$\mathbf{R}(\sigma, v) = \mathbf{Pr}\left[v \text{ is unblocked} \mid v \notin Y,\ Y(S_2(v)) = \sigma(S_2(v))\right],$$

where $S_2(v)$ are those vertices distance 2 from v and by "$v \notin Y$" we mean that v is not occupied under Y. Since G is triangle free, conditional on the configuration at v and $S_2(v)$ then the neighbors of v are independent in the Gibbs distribution. Substituting σ with Y in the above relation we have that

$$\mathbf{R}(Y, v) = \prod_{z \in N(v)} \mathbf{Pr}\left[z \notin Y \mid v \notin Y,\ Y(S_2(v))\right]. \tag{21}$$

In special cases of graphs, e.g., for Δ-regular trees. we can express the probability terms on the right-hand side (r.h.s.) of (21) in terms of $\mathbf{R}$-quantities. That is, we can extend (21) to the following system of recursive equations: With probability $1 - \exp(-\Omega(\Delta^{1/3}))$ we have

$$\mathbf{R}(Y, v) = \prod_{z \in N(v)} \left(1 - \frac{\lambda}{1+\lambda} \mathbf{R}(Y, z)\right) + O\left(\Delta^{-1/3}\right). \tag{22}$$

The above is not trivial to derive; in what follows we provide the technical details. For our purpose it turns out that $\mathbf{R}(X, \cdot)$ expressed as in (22) is an approximate version of $F()$ defined in (3). The error term $O(\Delta^{-1/3})$ in (22) is negligible. For understanding $\mathbf{R}(X, \cdot)$ qualitatively, this error term can be completely ignored.

To get some intuition of what is going to follow, consider the (BP system of) equations in (22). We show that a relation similar to (22) holds for the graph G. That is, we prove a loopy version of the equation. (See Lemma 12 for further details.) Furthermore, for proving the theorem we will show the following interesting result regarding the quantity $\mathbf{S}_Y(v)$ for every $v \in V$. With probability $\geq 1 - \exp(-\Omega(\Delta))$, it holds that

$$\left| \mathbf{S}_Y(v) - \sum_{z \in N(v)} \mathbf{R}(Y, z) \right| \leq \epsilon \Delta. \tag{23}$$

That is, we can approximate $\mathbf{S}_Y(v)$ by using quantities that arise from the loopy BP equations.

Still, getting a handle on $\mathbf{R}(Y,z)$ in (23) is a nontrivial task. We will argue that the loopy version of (22) we establish between $\mathbf{R}(Y,v)$ and $\mathbf{R}(Y,z)$ for $z \in N(v)$ is an approximate version of $F()$ and then we can apply Lemma 5 to deduce convergence (close) to the fixed point ω^*.

LEMMA 12. *For all $\gamma, \delta > 0$, there exists $\Delta_0, C > 0$, and for all graphs $G = (V,E)$ of maximum degree $\Delta \geq \Delta_0$ and girth ≥ 6 all $\lambda < (1-\delta)\lambda_c(\Delta)$ for all $v \in V$ the following is true:*

Let X be distributed as in μ. Then with probability $\geq 1 - \exp(-\Delta/C)$ it holds that

$$\left| \mathbf{R}(X,v) - \prod_{z \in N(v)} \left(1 - \frac{\lambda}{1+\lambda}\mathbf{R}(X,z)\right) \right| < \gamma. \tag{24}$$

Proof. Consider X distributed as in μ. Given some vertex $v \in V$, let $\mathcal{F}$ be the σ-algebra generated by the configuration of v and the vertices at distance ≥ 3 from v.

Note that $\lambda_c(\Delta) \sim e/\Delta$. So, for $\lambda < \lambda_c(\Delta)$ we have $\lambda = O(1/\Delta)$. Note, also, that $\mathbf{S}_X(v)$ is a function of the configuration at $S_2(v)$. Conditional on $\mathcal{F}$, for any $z, z' \in N(v)$ the configurations at $N(z)\backslash\{v\}$ and $N(z')\backslash\{v\}$ are independent of each other. That is, conditional on $\mathcal{F}$, the quantity $\mathbf{S}_X(v)$ is a sum of $|N(v)|$ many independent random variables in $\{0,1\}$. Then, applying Azuma's inequality (the Lipschitz constant is 1) we get that

$$\mathbf{Pr}\left[|\mathbb{E}\left[\mathbf{S}_X(v) \mid \mathcal{F}\right] - \mathbf{S}_X(v)| \leq \beta\Delta\right] \geq 1 - 2\exp\left(-\beta^2\Delta/2\right) \tag{25}$$

for any $\beta > 0$.

For $x \in \mathbb{R}_{\geq 0}$, let $f(x) = \exp(-\frac{\lambda}{1+\lambda}x)$. Since $\lambda \leq e/\Delta$ for $\Delta \geq \Delta_0$, then for $|\gamma| \leq (3e)^{-1}$ it holds that $f(x+\gamma\Delta) \leq 10\gamma$. Using these observations and (25) we get the following: for $0 < \beta < (3e)^{-1}$ it holds that

$$\mathbf{Pr}\left[|f(\mathbf{S}_X(v)) - f(\mathbb{E}\left[\mathbf{S}_X(v) \mid \mathcal{F}\right])| \leq 10\beta\right] \geq 1 - 2\exp\left(-\beta^2\Delta/2\right). \tag{26}$$

Recalling the definition of $\mathbf{R}(X,v)$, we have that

$$\begin{aligned}
\mathbf{R}(X,v) &= \prod_{z \in N(v)} \left(1 - \frac{\lambda}{1+\lambda}\mathbf{U}_{z,v}(X)\right) \\
&= \exp\left(-\frac{\lambda}{1+\lambda}\sum_{z \in N(v)} \mathbf{U}_{z,v}(X) + O\left(1/\Delta\right)\right) \\
&= f(\mathbf{S}_X(v)) + O\left(1/\Delta\right), \qquad (27)
\end{aligned}$$

where for the second equality we use the fact that $\lambda = O(1/\Delta)$ and that for $|x| < 1$ we have $1 + x = \exp(x + O(x^2))$; the last equality follows by noting that $f(\mathbf{S}_X(v)) \leq 1$.

We are now going to show that for every $z \in N(v)$ it holds that

$$|\mathbb{E}\left[\mathbf{U}_{z,v}(X) \mid \mathcal{F}\right] - \mathbf{R}(X,z)| \leq 2\lambda. \tag{28}$$

Before showing that (28) is indeed correct, let us show how we use it to get the lemma.

We have that

$$
\begin{aligned}
f(\mathbb{E}[\mathbf{S}_X(v) \mid \mathcal{F}]) &= \exp\left(-\frac{\lambda}{1+\lambda}\sum_{z\in N(v)} \mathbb{E}[\mathbf{U}_{z,v}(X_t) \mid \mathcal{F}]\right) \\
&= \exp\left(-\frac{\lambda}{1+\lambda}\sum_{z\in N(v)} \mathbf{R}(X,z)\right) + O(1/\Delta) \\
&= \prod_{z\in N(v)}\left(1-\frac{\lambda}{1+\lambda}\mathbf{R}(X,z)\right) + O(1/\Delta), \qquad (29)
\end{aligned}
$$

where in the first derivation we use linearity of expectation, in the second derivation we use (28) and the fact that $\lambda = O(1/\Delta)$, and in the third derivation we use the fact that $e^x = 1 + x + O(x^2)$ and the fact that $\lambda = (1/\Delta)$. The lemma follows by plugging (29) and (27) into (26) and taking sufficiently large Δ.

It remains to show (28). We first get an appropriate upper bound for $\mathbb{E}[\mathbf{U}_{z,v}(X) \mid \mathcal{F}]$. Using the fact that $\mathbf{U}_{z,w}(X) \le 1$ and $\mathbf{Pr}[z \in X|\mathcal{F}] \le \lambda$ we have that

$$
\begin{aligned}
\mathbb{E}[\mathbf{U}_{z,v}(X) \mid \mathcal{F}] &= \mathbb{E}[\mathbf{U}_{z,v}(X) \mid \mathcal{F}, z\in X\ ]\cdot \mathbf{Pr}[z\in X|\mathcal{F}] \\
&\quad + \mathbb{E}[\mathbf{U}_{z,v}(X) \mid \mathcal{F}, z\notin X]\cdot \mathbf{Pr}[z\notin X\ |\mathcal{F}] \\
&\le \mathbf{Pr}[z\in X|\mathcal{F}] + \mathbb{E}[\mathbf{U}_{z,v}(X) \mid \mathcal{F}, z\notin X] \\
&\le \lambda + \mathbb{E}[\mathbf{U}_{z,v}(X) \mid \mathcal{F}, z\notin X] \\
&= \lambda + \prod_{u\in N(z)\setminus\{v\}}\left(1-\frac{\lambda}{1+\lambda}\mathbf{U}_{u,z}(X)\right) \qquad (30) \\
&\le 2\lambda + \prod_{u\in N(z)}\left(1-\frac{\lambda}{1+\lambda}\mathbf{U}_{u,z}(X)\right) \\
&= 2\lambda + \mathbf{R}(X,z), \qquad (31)
\end{aligned}
$$

where (30) uses the fact that given $\mathcal{F}$ the values of $\mathbf{U}_{u,z}(X)$ for $u \in N(z)\setminus\{v\}$ are fully determined. Similarly, we get the lower bound:

$$
\begin{aligned}
&\mathbb{E}[\mathbf{U}_{z,v}(X) \mid \mathcal{F}] \\
&\quad = \mathbb{E}[\mathbf{U}_{z,v}(X) \mid \mathcal{F}, z\in X]\cdot \mathbf{Pr}[z\in X|\mathcal{F}] \\
&\qquad + \mathbb{E}[\mathbf{U}_{z,v}(X) \mid \mathcal{F}, z\notin X]\cdot \mathbf{Pr}[z\notin X|\mathcal{F}] \\
&\quad \ge \left(1-\frac{\lambda}{1+\lambda}\right)\mathbb{E}[\mathbf{U}_{z,v}(X) \mid \mathcal{F}, z\notin X] \qquad \left(\text{as } \mathbf{Pr}[z\notin X|\mathcal{F}] \ge 1-\tfrac{\lambda}{1+\lambda}\right) \\
&\quad \ge (1-2\lambda)\,\mathbb{E}[\mathbf{U}_{z,v}(X) \mid \mathcal{F}, z\notin X] \qquad \left(\text{as } \tfrac{\lambda}{1+\lambda} < 2\lambda\right) \\
&\quad \ge (1-2\lambda)\prod_{u\in N(z)\setminus\{w\}}\left(1-\frac{\lambda}{1+\lambda}\mathbf{U}_{u,z}(X)\right) \\
&\quad \ge (1-2\lambda)\prod_{u\in N(z)}\left(1-\frac{\lambda}{1+\lambda}\mathbf{U}_{u,z}(X)\right) \\
&\quad = (1-2\lambda)\mathbf{R}(X,z) \\
&\quad \ge \mathbf{R}(X,z) - 2\lambda, \qquad (32)
\end{aligned}
$$

where in the last inequality we use the fact that $\mathbf{R}(X,z) \le 1$.

From (31) and (32) we have proved (28), which completes the proof of the lemma. ∎

We will argue that (24) is an approximate version of $F(\cdot)$ and then we can apply Lemma 5 to deduce that $\mathbf{R}(X,\cdot)$ is an approximate version of the fixed point ω^*. In particular we have the following result.

LEMMA 13. *For every $\delta, \theta > 0$, there exists $\Delta_0 = \Delta_0(\delta,\theta)$ and $C > 0$ all $\lambda < (1-\delta)\lambda_c(\Delta)$, and for G of maximum degree Δ and girth ≥ 6, the following is true: Let X be distributed as the Gibbs distribution. For any $z \in V$, it holds that*

$$\mathbf{Pr}\left[|\mathbf{R}(X,z) - \omega^*(z)| \leq \theta\right] \geq 1 - \exp(-\Delta/C),$$

where ω^* is defined in Lemma 4.

Proof. Let $R = \lfloor \frac{50}{\delta}\log\theta^{-1} + \frac{100}{\delta} \rfloor$. For every integer $i \leq R$, we define

$$\beta_i := \max\left|\Psi(\mathbf{R}(X,x)) - \Psi(\omega^*(x))\right|,$$

where Ψ is defined in (5). The maximum is taken over all vertices $x \in B_i(w)$.

An elementary observation is that $\beta_i \leq 3$ for every $i \leq R$. To see why this holds, note that for any $z \in V$ and any independent sets σ, it holds that $e^{-e} \leq \mathbf{R}(\sigma,z), \omega^*(z) \leq 1$. Then we get $\beta_i \leq 3$ from (6).

We start by using the fact that $\beta_R \leq 3$. Then we show that with sufficiently large probability, if $\beta_{i+1} \geq \theta/5$, then $\beta_i \leq (1-\gamma)\beta_{i+1}$, where $0 < \gamma < 1$. Then the lemma follows by taking large R.

For any $i \leq R$, Lemma 12 and a simple union bound over the vertices in $B_i(w)$ imply that there exists a constant $C_0 = C_0(\theta,\delta) > 0$ such that with probability at least $1 - \exp(-\Delta/C_0)$ the following is true: For every vertex $x \in B_i(w)$ it holds that

$$\left|\mathbf{R}(X,x) - \exp\left(-\frac{\lambda}{1+\lambda}\sum_{z\in N(x)}\mathbf{R}(X,z)\right)\right| < \frac{\theta\delta}{40}. \tag{33}$$

Fix some $i \leq R$, $z \in B_i(w)$. From the definition of the quantity β_{i+1} we get the following: For any $x \in B_{i+1}(w)$ consider the quantity $\tilde{\omega}(x) = \mathbf{R}(X,x)$. We have that

$$D_{v,i+1}(\tilde{\omega},\omega^*) \leq \beta_{i+1}. \tag{34}$$

We will show that if (33) holds for $\mathbf{R}(X,z)$, where $z \in B_i(w)$, and $\beta_{i+1} \geq \theta/5$, then we have that

$$\left|\Psi\left(\mathbf{R}(X,z)\right) - \Psi\left(\omega^*(z)\right)\right| \leq (1-\delta/24)\beta_{i+1}.$$

For proving the above inequality, first note that if $\mathbf{R}(X,z)$ satisfies (33), then (6) implies that

$$\left|\Psi\left(\mathbf{R}(X,z)\right) - \Psi\left(\exp\left(-\frac{\lambda}{1+\lambda}\sum_{r\in N(z)}\mathbf{R}(X,r)\right)\right)\right| \leq \frac{\delta\theta}{12}. \tag{35}$$

Furthermore, we have that

$$
\begin{aligned}
&|\Psi(\mathbf{R}(X,z)) - \Psi(\omega^*(z))| \\
&\quad \le \frac{\delta\theta}{12} + \left|\Psi\left(\exp\left(-\frac{\lambda}{1+\lambda}\sum_{r\in N_z}\mathbf{R}(X,r)\right)\right) - \Psi(\omega^*(z))\right| \qquad \text{(from (35))} \\
&\quad \le \frac{\delta\theta}{12} + \left|\Psi\left(\prod_{r\in N(z)}\left(1-\frac{\lambda\mathbf{R}(X,r)}{1+\lambda}\right)\right) - \Psi(\omega^*(z))\right| + \\
&\quad + \left|\Psi\left(\prod_{r\in N(z)}\left(1-\frac{\lambda\mathbf{R}(X,r)}{1+\lambda}\right)\right) - \Psi\left(\exp\left(-\frac{\lambda}{1+\lambda}\sum_{r\in N(z)}\mathbf{R}(X,r)\right)\right)\right|, \qquad (36)
\end{aligned}
$$

where both inequalities follow from the triangle inequality.

From our assumption about λ and the fact that $\mathbf{R}(X,r) \in [e^{-e}, 1]$ for $r \in N(z)$, we have that

$$\left|\prod_{r\in N(z)}\left(1-\frac{\lambda\mathbf{R}(X,r)}{1+\lambda}\right) - \exp\left(-\lambda\sum_{r\in N(z)}\frac{\mathbf{R}(X,r)}{1+\lambda}\right)\right| \le \frac{10}{\Delta}.$$

The above inequality and (6) imply that

$$\left|\Psi\left(\prod_{r\in N(z)}\left(1-\frac{\lambda\mathbf{R}(X,r)}{1+\lambda}\right)\right) - \Psi\left(\exp\left(-\frac{\lambda}{1+\lambda}\sum_{r\in N(z)}\mathbf{R}(X,r)\right)\right)\right| \le \frac{30}{\Delta}.$$

Plugging the inequality above into (36) we get that

$$
\begin{aligned}
&|\Psi(\mathbf{R}(X,z)) - \Psi(\omega^*(z))| \\
&\quad \le \frac{\delta\theta}{12} + \frac{30}{\Delta} + \left|\Psi\left(\prod_{r\in N(z)}\left(1-\frac{\lambda\mathbf{R}(X,r)}{1+\lambda}\right)\right) - \Psi(\omega^*(z))\right| \\
&\quad \le \frac{\delta\theta}{12} + \frac{30}{\Delta} + \left|\Psi\left(\prod_{r\in N(z)}\left(\frac{1}{1+\lambda\mathbf{R}(X,r)}\right)\right) - \Psi(\omega^*(z))\right| \\
&\quad\quad + 3\left|\prod_{r\in N(z)}\left(\frac{1}{1+\lambda\mathbf{R}(X,r)}\right) - \prod_{r\in N(z)}\left(1-\frac{\lambda\mathbf{R}(X,r)}{1+\lambda}\right)\right| \qquad (37) \\
&\quad \le \delta\theta/12 + 60/\Delta + D_{v,i}(F(\tilde{\omega}), \omega^*), \qquad (38)
\end{aligned}
$$

where we derive (37) by applying the triangle inequality and (6); (38) follows by noting that for any $r \in N(z)$ we have $(\frac{1}{1+\lambda\mathbf{R}(X,r)}) - (1-\frac{\lambda\mathbf{R}(X,r)}{1+\lambda}) \le (e/\Delta)^2$, $|N(z)| \le \Delta$, and Δ is sufficiently large. Finally, in (38) we let $\tilde{\omega} \in [0,1]^V$ be such that $\tilde{\omega}(r) = \mathbf{R}(X,r)$ for $r \in V$ and F is defined in (3).

Since $\tilde{\omega}$ satisfies (34), Lemma 5 implies that

$$D_{v,i}(F(\tilde{\omega}), \omega^*) \le (1-\delta/6)\beta_{i+1}. \qquad (39)$$

Plugging (39) into (38) we get that

$$|\Psi(\mathbf{R}(X_s,z)) - \Psi(\omega^*(z))| \le \delta\theta/12 + 60/\Delta + (1-\delta/6)\beta_{i+1} \le (1-\delta/24)\beta_{i+1}, \qquad (40)$$

where the last inequality holds if we have $\beta_{i+1} \geq \theta/5$ and Δ is sufficiently large. Note that (40) holds provided that $\mathbf{R}(X, z)$ satisfies (33). The lemma follows by taking sufficiently large $R = R(\theta)$. ∎

Proof of Theorem 10. Let $\mathcal{F}$ be the σ-algebra generated by the configuration of v and the vertices at distance greater than 2 from x, i.e., $V \setminus B_2(v)$. Conditioning on $\mathcal{F}$, S_v is a sum of $|N(v)|$ many 0-1 independent random variables. From Azuma's inequality, for any fixed $\gamma > 0$, we have that

$$\mathbf{Pr}\left[|S_v - \mathbb{E}\left[S_v \mid \mathcal{F}\right]| > \gamma\Delta\right] \leq 2\exp\left(-\gamma^2\Delta/2\right). \tag{41}$$

Working as in the proof of Lemma 12 (i.e., for (31), (32)) we get the following: For each $z \in N(v)$ it holds that

$$|\mathbb{E}\left[\mathbf{U}_{z,v}(X) \mid \mathcal{F}\right] - \mathbf{R}(X, z)| \leq 10e^e\lambda.$$

Note that, given $\mathcal{F}$ the quantity $\mathbf{R}(X, z)$ is uniquely specified.

From the above we get that

$$\mathbb{E}\left[S_v \mid \mathcal{F}\right] = \sum_{z \in N(v)} \mathbb{E}\left[\mathbf{U}_{z,v}(X) \mid \mathcal{F}\right] = \sum_{z \in N(v)} \mathbf{R}(X, z) + \zeta, \tag{42}$$

where $|\zeta| \leq e^{5e}$. Furthermore, from Lemma 13 we have that for every $w \in V$ and every $\theta > 0$, there exists $C_0 > 0$ such that

$$\mathbf{Pr}\left[|\mathbf{R}(X, w) - \omega^*(w)| \leq \theta\right] \geq 1 - \exp\left(-\Delta/C_0\right). \tag{43}$$

From (43), (42), and a simple union bound we get the following: for every $\gamma' > 0$, there exists $C_1 > 0$ such that

$$\mathbf{Pr}\left[\left|\mathbb{E}\left[S_v \mid \mathcal{F}\right] - \sum_{z \in N(v)} \omega^*(z)\right| \leq \gamma'\Delta\right] \geq 1 - \exp\left(-\Delta/C_1\right). \tag{44}$$

The theorem follows by combining (41) and (44). ∎

5. Rapid mixing proof.

We begin with some basic notation. Consider a graph $G = (V, E)$. For some integer $r \geq 0$ and $v \in V$, let $B_r(v)$ be the set of vertices which are within distance r from v; we usually refer to $B_r(v)$ as the "ball" of radius r, centered at v. Let $S_r(v)$ the set of vertices at distance exactly r from v; we usually refer to $S_r(v)$ as the "sphere" of radius r, centered at v. Finally, let $N(v)$ denote the set of vertices which are adjacent to v.

5.1. Outline of the proof.

Theorem 11 tells us that after a burn-in period the Glauber dynamics locally behaves like the BP fixed points ω^* with high probability (whp). (In this discussion, we use the term whp to refer to events that occur with probability $\geq 1 - \exp(-\Omega(\Delta))$.) Meanwhile Theorem 9 says that there is an appropriate distance function $\mathcal{D}$ for which path coupling has contraction for pairs of states that behave as in ω^*. A snag in simply combining this pair of results and deducing rapid mixing is that when Δ is constant then there is still a constant fraction of the graph that does not behave like ω^* even in the stationary distribution, and the disagreements in our coupling proof may be biased toward this set. We follow the approach in [4] to overcome the obstacles that arise and complete the proof of Theorem 1.

The high level description of the proof of Theorem 1 is simple. The notions of local uniformity and the distance function $\mathcal{D}$, even though they are in the core of the rapid mixing analysis, do not appear in this level of description. We will discuss them a bit later in the exposition. At this stage we need to introduce the notion of a "heavy" vertex.

DEFINITION 14. *Let $G = (V, E)$ be a graph of maximum degree Δ and let σ be an independent set of G. For some $\rho > 0$, we say that σ is ρ-heavy for the vertex $v \in V$ if $|B_2(v) \cap \sigma| \geq \rho\Delta$ or $|B_1(v) \cap \sigma| \geq \rho\Delta/\log\Delta$.*

Heavy vertices are undesirable in that for a vertex v which is heavy, in order for its local neighborhood to attain the uniformity properties we need to first update most of its neighbors (or most of its grandchildren). This requires $\Omega(n \log \Delta)$ steps in which time disagreements spread far. In contrast for vertices v that are not heavy, and for which all vertices within some distance r from v are not heavy as described in the upcoming definition, we will prove that in $O(n)$ steps the local neighborhood of v attains the uniformity properties.

For our analysis we do not only care about some vertex v being heavy or not; we need to take into account for the heavy neighbors of v within some radius r around it, as well. More specifically, we introduce the following notions.

DEFINITION 15. *Let $G = (V, E)$ be a graph of maximum degree Δ. Let σ, τ be independent sets of G. Consider integer $r > 0$ and $v \in V$. If there is a vertex $w \in B_r(v)$ such that w is ρ-heavy, then σ is called (ρ, r)-bad at v. Otherwise, we say that σ is (ρ, r)-nice at v.*

Similarly, for σ, τ such that $\sigma(v) \neq \tau(v)$, we say that v is a (ρ, r)-bad disagreement if there exists a vertex $w \in B_r(v)$ such that either σ or τ is ρ-heavy at w. Otherwise, we say that v is a (ρ, r)-nice disagreement.

For the range of λ we consider here, a very useful observation about the Glauber dynamics (X_t) is that the bad vertices in the configuration X_t are very rare as long as $t = \Omega(n \log \Delta)$. In particular, in Lemma 21, we show that for the Glauber dynamics with fugacity $\lambda < \lambda_c$, after a burn-in period of length $\Omega(n \log \Delta)$, a vertex v becomes $(50, \Delta^{9/10})$-nice and remains nice for a period of length $ne^{\Omega(\Delta)}$ whp.

We prove Theorem 1 by employing path coupling. As it turns out, for the path coupling we need to focus on whether the disagreements we are dealing are bad or not. In our coupling analysis, bad disagreements have an increased tendency to create new ones. We need to use the fact that starting from a bad disagreement, after $\Omega(n \log \Delta)$ steps this disagreement is unlikely to remain bad.

Putting the above into a firmer basis, the coupling considers the pair of Markov chains (X_t) and (Y_t). We introduce a distance metric for the configurations of the chains. That is, we introduce a *weighted Hamming* distance β on the space of independent sets of the underlying graph G. For X_t, Y_t, we have that $\beta(X_t, Y_t)$ equals the sum of the Hamming distance between X_t, Y_t plus S times the number of $(200, r)$-bad disagreements of radius $r = 2\Delta^{3/5}$, where $S = \Delta^{3C'/\epsilon+1/2}$ for appropriate $C' > 0$ and $\epsilon > 0$.

At this point we need to remark that the distance β should not be confused with the metric $\mathcal{D}$ we introduced in section 1.3. The metric $\mathcal{D}$ is not used directly for the path coupling analysis but it is used later to derive some more technical results, i.e., in Lemma 17.

Rapid mixing follows by using path coupling to show contraction w.r.t. the metric β. We show contraction in a T-step coupling between (X_t) and (Y_t), where $T =$

$(C'/\epsilon)n(\log \Delta)$. In particular, given X_{iT} and Y_{iT}, for some integer $i \geq 0$, we show that there is a T-step coupling such that the expected distance of $X_{(i+1)T}$, $Y_{(i+1)T}$ is much smaller than $\beta(X_{iT}, Y_{iT})$, i.e.,

$$\mathbb{E}\left[\beta\left(X_{(i+1)T}, Y_{(i+1)T}\right) \mid X_{iT}, Y_{iT}\right] \leq \frac{2}{\sqrt{\Delta}}\beta\left(X_{iT}, Y_{iT}\right). \tag{45}$$

Sketching the proof of (45) we have the following: Consider X_{iT}, Y_{iT} with ℓ disagreements out of which h are $(200, 2\Delta^{3/5})$-bad. Then we have that

$$\beta\left(X_{(i+1)T}, Y_{(i+1)T}\right) = \ell' + S \cdot h',$$

where ℓ' is the number of disagreements between $X_{(i+1)T}, Y_{(i+1)T}$ and h' is the number of the disagreements which are $(200, 2\Delta^{3/5})$-bad. Then, (45) follows by bounding appropriately $\mathbb{E}[\ell']$ and $\mathbb{E}[h']$.

We apply path coupling to (X_{iT}, Y_{iT}). Consider the interpolating sequence $Z_0, \ldots, Z_\ell$, such that $Z_0 = X_{iT}$, $Z_\ell = Y_{iT}$ and, for $0 \leq j < \ell$, the pair Z_j, Z_{j+1} differ on the assignment of a single vertex, say, vertex w_j. We couple Z_j and Z_{j+1} and let Z'_j, Z'_{j+1} be the pair of configurations we get after T steps.

There is a straightforward argument that if vertex w_j is a $(200, 2\Delta^{3/5})$-nice disagreement for X_{iT} and Y_{iT}, then we can have the interpolating sequence such that w_j is a $(200, 2\Delta^{3/5})$-nice disagreement for Z_j and Z_{j+1}.

First, we get an upper bound on the expected number of disagreements in the pair Z'_j, Z'_{j+1}. Note that some of these disagreements are nice and some are $(200, 2\Delta^{3/5})$-bad. We are going to bound the expectation of these two kinds of disagreement, separately. Once we get the expected number of disagreements (for both kinds) for each pair Z'_j, Z'_{j+1}, a standard argument from path coupling gives $\mathbb{E}[h'], \mathbb{E}[\ell']$.

As far as $\mathbb{E}[h']$ is concerned we use the following lemma.

LEMMA 16. *For $\delta > 0$, $0 < \epsilon < 1$ and $C > 10$ let $\Delta \geq \Delta_0$. Consider a graph $G = (V, E)$ of maximum degree Δ and let $\lambda \leq (1-\delta)\lambda_c(\Delta)$. Let $(X_t), (Y_t)$ be the Glauber dynamics on the hard-core model with fugacity λ and underlying graphs G. Assume that the two chains are maximally coupled. Then, the following is true:*

Assume X_0, Y_0 to be such that $X_0 \oplus Y_0 = \{v^*\}$ and $T = C'n/\epsilon$. Then it holds that

1. $\mathbb{E}\left[|X_{T\log\Delta} \oplus Y_{T\log\Delta}|\right] \leq \Delta^{3C'/\epsilon}$,
2. *let $S_{T\log\Delta}$ denote the set of disagreements of $(X_{T\log\Delta}, Y_{T\log\Delta})$ that are $(200, r)$-bad for radius $2\Delta^{3/5}$. Then $\mathbb{E}\left[|S_{T\log\Delta}|\right] \leq \exp(-\sqrt{\Delta})$.*

For the bounds in Lemma 16 we do not need to use any uniformity arguments. Mainly we use worst-case assumptions regarding the generation of disagreements. Lemma 16 follows as a corollary from Lemma 23 which we present and prove in section 5.4.

From Lemma 16.2 we have that there is a coupling such that the expected number of disagreements between Z'_j, Z'_{j+1} which are $(200, 2\Delta^{3/5})$-bad is $\leq \exp(-\sqrt{\Delta})$. Path coupling then implies that

$$\mathbb{E}[h'] \leq \ell \exp(-\sqrt{\Delta}).$$

Furthermore, since the number of $(200, 2\Delta^{3/5})$-bad disagreements between X_{iT}, Y_{iT} is assumed to be h, there are at most h pairs Z_j, Z_{j+1} such that the disagreement is $(200, 2\Delta^{3/5})$-bad. Lemma 16.1 implies that there is a coupling such that the expected number of disagreements generated by such a pair is $\leq \Delta^{3C'/\epsilon}$.

For each of the rest of the pairs Z_j, Z_{j+1}, i.e., those pairs whose disagreement is $(200, 2\Delta^{3/5})$-nice we use the following result.

LEMMA 17. *Let $C' > 10$, $\epsilon > 0$ and $\Delta_0 = \Delta_0(\epsilon)$. For any graph $G = (V, E)$ on n vertices, of maximum degree $\Delta > \Delta_0$, girth $g \geq 7$, and for $\lambda \leq (1-\delta)\lambda_c(\Delta)$ the following is true.*

Let $(X_t), (Y_t)$ be the Glauber dynamics on the hard-core model with fugacity λ and underlying graphs G. Let X_0, Y_0 be independent sets which disagree on a single vertex v^* that is $(400, R)$-nice for radius $R = 2\Delta^{3/5}$. For $T = (C'/\epsilon)n \log \Delta$, we have that

$$\mathbb{E}\left[|X_T \oplus Y_T|\right] \leq 1/\sqrt{\Delta}.$$

For each of the pairs Z_j, Z_{j+1} whose disagreement is $(200, 2\Delta^{3/5})$-nice, Lemma 17 implies that we have contraction. That is, the expected number of disagreements is $1/\sqrt{\Delta}$. Putting together the bounds for the nice and bad pairs we have that

$$\mathbb{E}\left[\ell'\right] \leq (\ell - h)/\sqrt{\Delta} + h\Delta^{3C'/\epsilon}.$$

We require that the maximum degree Δ is large enough so that the quantity S, in the definition of the metric β, satisfies $S \leq \exp(\sqrt{\Delta})/\sqrt{\Delta}$. With this requirement for Δ and the above bounds for $\mathbb{E}\left[\ell'\right]$ and $\mathbb{E}\left[h'\right]$, it is a matter of elementary calculations to verify that (45) indeed holds.

To summarize all above, we have the following: for the bad disagreements we should expect a "bad behavior" in terms of the new disagreements they create. That is, if w_j is bad, then the expected number of disagreements after a T-step coupling of Z_j, Z_{j+1} is at most $\Delta^{3C'/\epsilon}$, as specified from Lemma 16. On the other hand, the bad disagreements tend to be rare after a T-step coupling, i.e., regardless of whether w_j is heavy or not, the expected number of new heavy disagreements that are created is at most $\exp(-\sqrt{\Delta})$; this also follows from Lemma 16. That is, even though the heavy disagreements create a lot of disagreements, they tend to be rare after a while and their expected contribution is minuscule. The main load for proving rapid mixing comes from the nice disagreements. For them, we use Lemma 17, which essentially uses the uniformity result in Theorem 18.

Contraction using uniformity. We use the notion of local uniformity to prove Lemma 17. In the remainder of this section, we sketch the basic results we prove to derive the lemma.

We use the following, more technical version of the uniformity result contained in Theorem 11.

THEOREM 18. *For all $\delta, \epsilon > 0$, let $\Delta_0 = \Delta_0(\delta, \epsilon), C = C(\delta, \epsilon)$. For graph $G = (V, E)$ of maximum degree $\Delta \geq \Delta_0$ and girth ≥ 7, all $\lambda < (1-\delta)\lambda_c(\Delta)$, let (X_t) be the continuous (or discrete) time Glauber dynamics on the hard-core model. If X_0 is $(400, R)$-nice at $v \in V$ for radius $R = R(\delta, \epsilon) > 1$, it holds that*

$$\mathbf{Pr}\left[(\forall t \in \mathcal{I}) \quad \mathbf{W}_{X_t}(v) < \sum_{z \in N(v)} \omega^*(z)\Phi(z) + \epsilon\Delta\right] \geq 1 - \exp\left(-\Delta/C\right), \tag{46}$$

where the time interval $\mathcal{I} = [Cn, n\exp(\Delta/C)]$.

Note that the only difference between Theorem 11 and the one above is that the later assumes that X_0 is $(400, R)$-nice at v for radius $R = R(\delta, \epsilon) > 1$ and only $O(n)$ steps are required to attain the local uniformity properties. In contrast, Theorem 11 does not make any assumption about X_0, and hence $O(n \log \Delta)$ steps are required. The requirement for $O(n \log \Delta)$ steps to get uniformity comes from the fact that the vertex v for which we want to establish uniformity, at X_0, is heavy. Then, $O(n \log \Delta)$ steps allow all for the neighbors of v to be updated at least once, whp.

The proof of Theorem 18 is quite technical and goes beyond the discussion of this section. For this reason, the presentation of the proof is deferred to section 10.

We also need to use the following Theorem 19, which shows that for (X_t) and (Y_t) such that X_0 and Y_0 specify only a single, nice disagreement, there is an $O(n)$-step coupling where the expected Hamming distance decreases.

THEOREM 19. *Let $C' > 10$, $\delta, \epsilon > 0$, let $\Delta_0 = \Delta_0(\epsilon, \delta)$, and let $\lambda < (1-\delta)\lambda_c(\Delta)$. For any graph $G = (V, E)$ on n vertices and maximum degree $\Delta > \Delta_0$ and girth $g \geq 7$ the following holds:*

Let $(X_t), (Y_t)$ be the Glauber dynamics on the hard-core model with fugacity λ and underlying graphs G. Suppose that X_0, Y_0 differ only at v^*, while v^* is $(400, R)$-nice for R, where $\Delta^{3/5} \leq R \leq 2\Delta^{3/5}$. For $T_m = C'n/\epsilon$ we have that

1. $\mathbb{E}\left[|X_{T_m} \oplus Y_{T_m}|\right] \leq 1/3$,
2. *let $\mathcal{Z}$ denote the event that there exists a $(200, R')$-bad disagreement for $R' = R - 2\sqrt{\Delta}$ at time T_m. Then it holds that*

$$\mathbf{Pr}\left[\mathcal{Z}\right] \leq 2\exp(-2\sqrt{\Delta}).$$

The main argument for proving Theorem 19 is as follows: We let the two chains run for some $\Theta(n)$ steps, such that v^* and $B_{\sqrt{\Delta}}(v^*)$ get uniformity. For this argument we combine Theorem 18 and a simple union bound over the vertices in $B_{\sqrt{\Delta}}(v^*)$. Note that the period we wait for $B_{\sqrt{\Delta}}(v^*)$ to get uniformity is only a small fraction of T_m, the period we consider for the coupling.

During this initial period there is not much control on the number of disagreements that are created between the coupled chains. That is, we have to make worst case assumptions on how the new disagreements are generated. As it turns out there is an ever increasing number of new disagreements as we allow the two chains to evolve.

Despite this lack of control on the new disagreements, a key observation in our argument is that whp they are confined within the ball $B_{\sqrt{\Delta}}(v^*)$. Furthermore, whp the vertices in this ball get uniformity. Then we have contraction in the path coupling condition (by applying Theorem 9), and hence after $O(n)$ further steps the expected Hamming distance is small (by Theorem 19).

Note that Lemma 17 considers time interval $\Theta(n \log \Delta)$ as opposed to Theorem 19, which considers intervals of length $O(n)$. We get Lemma 17 by splitting the $\Theta(n \log \Delta)$ interval into epochs of length $\Theta(n)$, where for each epoch we apply Theorem 19. Then, the technical challenge in proving Lemma 17 amounts to arguing that certain, relatively rare, undesirable events, like the generation of bad disagreements, do not have much influence on the creation of new disagreements.

5.2. Proof of Theorem 1. For the purposes of path coupling for every pair of independent sets X, Y we consider shortest paths between X and Y along neighboring independent sets. That is, $X = Z_0 \sim Z_1 \sim \cdots \sim Z_\ell = Y$. This sequence $Z_1, \ldots, Z_\ell$ we call interpolated independent sets for X and Y and for any $i = 0, \ldots, \ell-1$ it holds that $|Z_i \oplus Z_{i+1}| = 1$.

Since the distance between neighboring configurations depends only on the vertex on which they disagree, we can move from X to Y by first removing the vertices in

$X \setminus Y$ and then adding the vertices in $Y \setminus X$. A key aspect of the above definitions is that the niceness is inherited by interpolated independent sets.

Observation 20. If X, Y are independent sets, neither of which is ρ-heavy at vertex v, then no interpolated independent set is ρ-heavy at v. Likewise, if v is (ρ, r)-nice, then in every interpolated independent sets v is (ρ, r)-nice.

The above follows from the fact that if independent sets σ, τ are such that $\sigma \subseteq \tau$, then a vertex v that is not ρ-heavy in τ is not ρ-heavy σ, as well.

Proof of Theorem 1. The proof of the theorem is very similar to the proof of [4, Theorem 1].

As we saw in the proof sketch, we define a *weighted Hamming* distance β on the space of independent sets. For X_t, Y_t, we have that $\beta(X_t, Y_t)$ equals the sum of the Hamming distance between X_t, Y_t plus S times the number of $(200, r)$-bad disagreements of radius $r = 2\Delta^{3/5}$, where $S = \Delta^{3C'/\epsilon + 1/2}$ for sufficiently large $C' > 0$. The theorem will follow by using path coupling and showing contraction w.r.t. the metric β.

For showing Theorem 1 we need to apply path coupling to an arbitrary pair of initial configurations. This implies that we need to deal with pairs whose disagreement is heavy. We introduce the metric β, mainly, to deal with the cases that the coupling of a pair of configurations starts from a heavy disagreement.

We require that the maximum degree Δ is large enough so that the above quantity S satisfies $S \leq \exp(\sqrt{\Delta})/\sqrt{\Delta}$.

In the following claim we show that we have contraction w.r.t. to the metric β. In particular, given X_{iT} and Y_{iT}, for some integer $i \geq 0$, there is a T-step coupling such that the expected distance of $X_{(i+1)T}, Y_{(i+1)T}$ is much smaller than $\beta(X_{iT}, Y_{iT})$, where $T = C'n(\log \Delta)/\epsilon$.

In particular we are going to show (45), which we restate here. For any $i \geq 0$, we have that

$$\mathbb{E}\left[\beta\left(X_{(i+1)T}, Y_{(i+1)T}\right) \mid X_{iT}, Y_{iT}\right] \leq \frac{2}{\sqrt{\Delta}} \beta\left(X_{iT}, Y_{iT}\right). \tag{45}$$

Before showing that (45) is true, let us show how it implies Theorem 1.

Using induction and (45) we obtain the following for $T = C'n(\log \Delta)/\epsilon$:

$$\mathbb{E}\left[\beta\left(X_{iT}, Y_{iT}\right)\right] \leq \left(\frac{2}{\sqrt{\Delta}}\right)^i \times \beta_{\max}, \tag{47}$$

where $\beta_{\max}$ is the maximum possible distance between two configurations, i.e., $\beta_{\max} = (S+1)n$. Choosing $i^* = \frac{C_1 \log(n/\delta)}{C' \log \Delta}$, where C_1 is sufficiently larger than C', from (47) and Markov's inequality we get that

$$\mathbf{Pr}\left[X_{i^*T} \neq Y_{i^*T}\right] \leq \mathbb{E}\left[\beta\left(X_{i^*T}, Y_{i^*T}\right)\right] \leq \delta.$$

The above inequality implies that $T_{\text{mix}}(\delta) \leq i^* T = (C_1/\epsilon) \log(n/\delta)$, which implies the theorem.

It remains to show (45). For this, we follow the steps we described in section 5.1. Consider X_{iT}, Y_{iT} with ℓ disagreements out of which h are $(200, 2\Delta^{3/5})$-bad. Then we have that

$$\beta\left(X_{(i+1)T}, Y_{(i+1)T}\right) = \ell' + S \cdot h',$$

where ℓ' is the number of disagreements between $X_{(i+1)T}, Y_{(i+1)T}$ and h' is the number of these disagreements which are $(200, 2\Delta^{3/5})$-bad.

Equation (45) follows by bounding appropriately $\mathbb{E}[\ell']$ and $\mathbb{E}[h']$. For this, we apply path coupling to (X_{iT}, Y_{iT}). Consider the interpolating sequence $Z_0, \ldots, Z_\ell$, such that $Z_0 = X_{iT}$, $Z_\ell = Y_{iT}$, and, for $0 \le j < \ell$, the pair Z_j, Z_{j+1} differ on the assignment of a single vertex, say, vertex w_j. We couple Z_j and Z_{j+1} and let Z'_j, Z'_{j+1} be the pair of configurations we get after T steps. First, we get an upper bound on the expected number of disagreements as well as the expected number of $(200, 2\Delta^{3/5})$-bad disagreements in the pair Z'_j, Z'_{j+1}. Then, path coupling implies the desired bounds for $\mathbb{E}[h'], \mathbb{E}[\ell']$.

As far as $\mathbb{E}[h']$ is concerned, from Lemma 16.2 we have that there is a coupling such that the expected number of disagreements between Z'_j, Z'_{j+1} which are $(200, 2\Delta^{3/5})$-bad is $\le \exp(-\sqrt{\Delta})$ and hence $\mathbb{E}[h'] \le \ell \exp(-\sqrt{\Delta})$.

Since the number of $(200, 2\Delta^{3/5})$-bad disagreements between X_{iT}, Y_{iT} is assumed to be h, there are at most h pairs Z_j, Z_{j+1} such that the disagreement is $(200, 2\Delta^{3/5})$-bad. Lemma 16.1 implies that there is a coupling such that the expected number of disagreements generated by the pair is $\le \Delta^{3C'/\epsilon}$. For each of the rest of the pairs Z_j, Z_{j+1} (namely, those pairs whose disagreement is $(200, 2\Delta^{3/5})$-nice), Lemma 17 implies that the expected number of disagreements is $1/\sqrt{\Delta}$. Putting together the bounds for the nice and bad pairs we have that $\mathbb{E}[\ell'] \le (\ell - h)/\sqrt{\Delta} + h\Delta^{3C'/\epsilon}$.

With the previous bounds for $\mathbb{E}[\ell']$ and $\mathbb{E}[h']$, it is a matter of elementary calculations to verify that (45) indeed holds for $C' > 0$ sufficiently large.

This concludes the proof of Theorem 1. ∎

5.3. Burn-in. The following results that we provide in this section are standard and we use them not only for proving rapid mixing of Glauber dynamics but in other places, i.e., for our uniformity results. For this reason we consider both continuous and discrete time Glauber dynamics. In the continuous time Glauber dynamics, the spin of each vertex is updated according to an independent Poisson clock with rate $1/n$.

The following lemma states that (X_t) requires $O(n \log \Delta)$ to burn-in, regardless of X_0.

LEMMA 21. *For $\delta > 0$ let $\Delta \ge \Delta_0(\delta)$ and $C_b = C_b(\delta)$. Consider a graph $G = (V, E)$ of maximum degree Δ. Also, let $\lambda \le (1-\delta)\lambda_c(\Delta)$.*

Let (X_t) be the continuous (or discrete) time Glauber dynamics on the hard-core model with fugacity λ and underlying graph G. Consider $v \in V$ and let $\mathcal{C}_t$ be the event that X_t is $(50, r)$-nice at v for radius $r = \Delta^{9/10}$. Then, for $\mathcal{I} = [10n \log \Delta, n \exp(\Delta/C_b)]$ it holds that

$$\mathbf{Pr}\left[\bigcap_{t \in \mathcal{I}} \mathcal{C}_t\right] \ge 1 - \exp\left(-\Delta/C_b\right).$$

Proof. For now, consider the continuous time version of (X_t). Recall that for X_t, the vertex u is not ρ-heavy if both of the following conditions hold:

1. $|X_t \cap B_2(u)| \le \rho\Delta$.
2. $|X_t \cap N(u)| \le \rho\Delta/\log\Delta$.

First we consider a fixed time $t \in \mathcal{I}$. Let $c = t/n$. Note that $c = c(\Delta) \ge 10\log\Delta$. We are going to show that there exists $C' > 0$ such that

$$\mathbf{Pr}[\mathcal{C}_t] \ge 1 - \exp\left(-\Delta/C'\right). \tag{48}$$

Fix some vertex $u \in B_r(v)$. Let N_0 be the set of vertices in $B_2(u) \cap X_0$ which are not updated during the time period $(0, t]$. That is, for $z \in N_0$ it holds that $X_0(z) =$

$X_t(z)$. Each vertex $z \in B_2(u) \cap X_0$ belongs to N_0 with probability $\exp(-t/n) = e^{-c}$, independently of the other vertices. Since $|B_2(u) \cap X_0| \leq \Delta^2$, it is elementary that the distribution of $|N_0|$ is dominated by $\mathcal{B}(\Delta^2, e^{-c})$, i.e., the binomial with parameters Δ^2 and e^{-c}.

Using Chernoff's bounds we get the following: for $c > 10 \log \Delta$ it holds that

$$\mathbf{Pr}\left[N_0 > \Delta/10\right] \leq \exp\left(-\Delta/10\right). \tag{49}$$

Additionally, let $N_1 \subseteq B_2(u)$ contain every vertex u which is updated at least once during the period $(0, t]$. Each vertex $z \in N_1$, which is last updated prior to t at time $s \leq t$, becomes occupied during the update at time s with probability at most $\frac{\lambda}{1+\lambda}$, regardless of $X_s(N(z))$. Then, it is direct that $|X_t \cap N_1|$ is dominated by $\mathcal{B}(N_1, \frac{\lambda}{1+\lambda})$.

Noting that $|N_1| \leq |B_2(u)| \leq \Delta^2$ and $\frac{\lambda}{1+\lambda} < 2e/\Delta$, for $\Delta > \Delta_0$ Chernoff's bound implies that

$$\mathbf{Pr}\left[|N_1 \cap X_t| \geq 15e\Delta\right] \leq \exp\left(-15e\Delta\right). \tag{50}$$

From (49), (50), and a simple union bound, we get that

$$\mathbf{Pr}\left[|X_t \cap B_2(u)| > 42\Delta\right] \leq \exp\left(-\Delta/20\right). \tag{51}$$

Using exactly the same arguments, we also get that

$$\mathbf{Pr}\left[|X_t \cap N(u)| > 42\Delta/\log \Delta\right] \leq \exp\left(-\Delta/20\right). \tag{52}$$

Note that X_0 could be such that $|N(u) \cap X_0| = \alpha\Delta$ for some fixed $\alpha > 0$. So as to get $|X_t \cap N(u)| \leq 42\Delta/\log\Delta$ with large probability, we have to ensure that with large probability all the vertices in $N(v)$ are updated at least once. For this reason the burn-in requires at least $10n \log \Delta$ steps.

From (51) and (52) we get the following: For any $\rho > 50$ it holds that

$$\mathbf{Pr}\left[X_t(u) \text{ is not } \rho\text{-heavy}\right] \leq \exp\left(-\Delta/25\right). \tag{53}$$

Then (48) follows by taking a union bound over all the at most Δ^r vertices in $B_r(v)$. In particular, for $r = \Delta^{9/10}$ and sufficiently large Δ, there exists $C > 0$ such that

$$\mathbf{Pr}\left[\mathcal{C}_t\right] \leq \Delta^r \exp\left(-\Delta/25\right) \leq \exp\left(-\Delta/30\right).$$

The above implies that (48) is indeed true but only for a specific time step $t \in \mathcal{I}$. Now we use a covering argument to deduce the above for the whole interval $\mathcal{I}$.

For sufficiently small $\gamma > 0$, independent of Δ, consider a partition of the time interval $\mathcal{I}$ into subintervals each of length $\frac{\gamma^2}{\Delta}n$, (where the last part can be shorter). We let $T(j)$ be the jth part in the partition.

Each $z \in B_2(w)$ is updated at least once during the time period $T(j)$ with probability less than $2\frac{\gamma^2}{\Delta}$, independently of the other vertices. Note that $|B_2(w)| \leq \Delta^2$. Clearly, the number of vertices in $B_2(v)$ which are updated during $T(j)$ is dominated by $\mathcal{B}(\Delta^2, 2\gamma^2/\Delta)$. Chernoff bounds imply that with probability at least $1 - \exp(-20\Delta\gamma^2)$, the number of vertices in $B_2(w)$ which are updated during the interval $T(j)$ is at most $20\gamma^2\Delta$. Furthermore, changing any $20\Delta\gamma^2$ variables in $B_2(w)$ can only make the independent set heavier by at most $20\Delta\gamma^2$.

Similarly, we get that with probability at least $1 - \exp(-\gamma\Delta)$, the number of vertices in $N(v)$ which are updated during the interval $T(j)$ is at most $\gamma\Delta/\log\Delta$. The change of at most $\gamma\Delta/\log\Delta$ neighbors of v does not change the weight of its neighborhood by more than $\gamma\Delta/\log\Delta$.

From the above arguments we get that the following: We can choose sufficiently large $C_b > 0$ such that for $j \in \{1, 2, \ldots, \lceil \Delta/(\gamma^2)\exp(\Delta/C_b)\rceil\}$ it holds that

$$\mathbf{Pr}\left[\cap_{t\in T(j)}\mathcal{C}_t\right] \geq 1 - \exp\left(-100\Delta/C_b\right).$$

The result for continuous time follows by taking a union bound over all the $\lceil \Delta/(\gamma^2)\exp(\Delta/C_b)\rceil$ many subintervals of $\mathcal{I}$.

For the discrete time case the arguments are very similar. The only extra ingredient we need is that, now, the updates of the vertices are negatively associated. The concentration inequalities above still hold since Chernoff's bounds hold for negatively associated random variables, e.g, see [3, Proposition 7]. The lemma follows. ∎

The following lemma states that if (X_t) start from a not so heavy state it only requires $O(n)$ steps to burn in.

LEMMA 22. *For $\delta > 0$, let $\Delta \geq \Delta_0(\delta)$ and $C_b = C_b(\delta)$. Consider a graph $G = (V, E)$ of maximum degree Δ. Also, let $\lambda \leq (1-\delta)\lambda_c(\Delta)$.*

Let (X_t) be the continuous (or discrete) time Glauber dynamics on the hard-core model with fugacity λ and underlying graph G. Consider $v \in V$ and let $\mathcal{C}_t$ be the event that X_t is $(50, R)$-nice at v for radius $R \leq \Delta^{9/10}$. Assume that X_0 is $(400, R)$-nice at v. Then, for $\mathcal{I} = [C_b n, n\exp(\Delta/(C_b\log\Delta))]$ we have that

$$\mathbf{Pr}\left[\cap_{t\in\mathcal{I}}\mathcal{C}_t\right] \geq 1 - \exp\left(-\Delta/(C_b\log\Delta)\right).$$

The proof of Lemma 22 is almost identical to the proof of Lemma 21, and for this reason we omit it.

In light of Lemma 21, Theorem 11 follows as a corollary from Theorem 18.

5.4. Expected Hamming distance for worst-case pair. The following lemma considers a worst case pair of neighboring independent sets. It states some upper bounds on the Hamming distance after Cn and $Cn\log\Delta$ steps of the coupling.

LEMMA 23. *For $\delta > 0$, $0 < \epsilon < 1$, and $C > 10$ let $\Delta \geq \Delta_0$. Consider a graph $G = (V, E)$ of maximum degree Δ and let $\lambda \leq (1-\delta)\lambda_c(\Delta)$. Let $(X_t), (Y_t)$ be the Glauber dynamics on the hard-core model with fugacity λ and underlying graphs G. Assume that the two chains are maximally coupled. Then, the following is true:*

Assume X_0, Y_0 to be such that $X_0 \oplus Y_0 = \{v^*\}$ and $T = Cn/\epsilon$. Then it holds that

1. $\mathbb{E}\left[|X_T \oplus Y_T|\right] \leq \exp(3C/\epsilon)$.
2. $\mathbb{E}\left[|X_{T\log\Delta} \oplus Y_{T\log\Delta}|\right] \leq \Delta^{3C/\epsilon}$.
3. *Let $\mathcal{E}_T$ be the event that at some time $t \leq T$, $|X_t \oplus Y_t| > \Delta^{2/3}$. Then*

$$\mathbb{E}\left[|X_T \oplus Y_T| \cdot \mathbf{1}\{\mathcal{E}_T\}\right] < \exp(-\sqrt{\Delta}).$$

4. *Let $S_{T\log\Delta}$ denote the set of disagreements of $(X_{T\log\Delta}, Y_{T\log\Delta})$ that are $(200, r)$-bad for radius $2\Delta^{3/5}$. Then $\mathbb{E}\left[|S_{T\log\Delta}|\right] \leq \exp(-\sqrt{\Delta})$.*

Note that Lemmas 23.2 and 23.4 correspond to Lemma 16. That is, Lemma 16 is contained in the statement of Lemma 23. We prove each of the four statements of Lemma 23 in turn.

Proof of Lemmas 23.1 *and* 23.2. The treatments for both cases are very similar. Note that each vertex can only become disagreeing at time step t if it is updated at time t and it is next to a vertex which is also disagreeing. Furthermore, for such vertex the probability to become disagreeing is at most e/Δ. Using the observations and noting that each disagreeing vertex has at most Δ nondisagreeing neighbors we get the following: The expected number of disagreements at each time step increases by a factor which is at most $(1+\Delta\frac{e}{n\Delta}) \le \exp(3/n)$.

By using induction, it is straightforward that for any $t \ge 0$ it holds that

$$\mathbb{E}\left[X_t \oplus Y_t\right] \le \exp\left(3t/n\right). \tag{54}$$

Then, Statement 1 follows by plugging into (54) the time $t = Cn/\epsilon$. Statement 2 follows by plugging into (54) the time $t = T\log\Delta$. ∎

Proof of Lemma 23.3. Recall that for any X_t, Y_t, we have that $D_t = \{w : X_t \neq Y_t\}$, while letting

$$D_{\le t} = \bigcup\nolimits_{t' \le t} D_{t'}.$$

Also, let $H_{\le t} = |D_{\le t}|$. We prove that for any integer $1 \le \ell \le n$, for $T = Cn/\epsilon$, it holds that

$$\mathbf{Pr}\left[H_{\le T} \ge \ell\right] \le \exp\left(-(\ell-1)e^{-6C/\epsilon}\right). \tag{55}$$

For $1 \le i \le \ell$, let t_i be the time at which the ith disagreement is generated (possibly counting the same vertex set multiple times). Denote $t_0 = 0$. Let $\eta_i := t_i - t_{i-1}$ be the waiting time for the formation of the ith disagreement. Since we assumed that $X_0 \oplus Y_0 = \{v^*\}$, we have that $\eta_1 = 0$. For $i \ge 1$, conditioned on the evolution at all times in $[0, t_i]$, the distribution of η_{i+1} stochastically dominates a geometric distribution with success probability ρ_i and range $\{1, 2, \ldots\}$, where

$$\rho_i = \frac{e \cdot \min\{i\Delta, n-i\}}{n\Delta}.$$

This is because at all times prior to t_{i+1} we have $H_t \le i$, while the sets $H_{\le t}$ increases with probability at most ρ_i at each time step, regardless of the history. The quantity $\min\{i\Delta, n-i\}$ in the numerator in the expression for ρ_i is an upper bound on the number of vertices that are nondisagreeing neighbors of the disagreeing vertices. The quantity $e/(n\Delta)$ is an upper bound for the probability of a neighbor of a disagreement to be chosen and become a disagreement itself.

Hence, $\eta_1 + \cdots + \eta_\ell$ stochastically dominates the sum of independent geometrically distributed random variables with success probability $\rho_1, \ldots, \rho_{\ell-1}$. For any real $x \ge 0$ it holds that

$$\mathbf{Pr}\left[\eta_{i+1} \ge x\right] \ge (1-\rho_i)^{\lceil x \rceil - 1} \ge \exp\left[-\frac{\rho_i}{1-\rho_i}x\right] \ge e^{2\rho_i x}.$$

In the above series of inequalities we used that $1 - x > \exp(-\frac{x}{1-x})$ for $0 < x < 1$ and $\rho_i < 1/3$.

The above inequality implies that $\eta_1 + \cdots + \eta_\ell$ dominates the sum of exponential random variables with parameters $2\rho_1, 2\rho_2, \ldots, 2\rho_{\ell-1}$. Since $\rho_i \le i\rho$, where $\rho = \frac{e}{n}$, we have that $\eta_1 + \cdots + \eta_\ell$ stochastically dominates the sum of exponential random variables $\zeta_1, \zeta_2, \ldots, \zeta_{\ell-1}$ with parameters $2\rho, 4\rho, \ldots, 2(\ell-1)\rho$, respectively.

Consider the problem of collecting $\ell - 1$ coupons, assuming that each coupon is generated by a Poisson process with rate 2ρ. The time interval between collecting the ith coupon and the $i+1$'st coupon is exponentially distributed with rate $2(\ell-1-i)\rho$. Hence the time to collect all $\ell-1$ coupons has the same distribution as $\zeta_1 + \zeta_2 + \cdots + \zeta_{\ell-1}$. But the event that the total delay is less than T is nothing but the intersection of the (independent) events that all coupons are generated in the time interval $[0, T]$. The probability of this event is

$$(1 - \exp^{-2T\rho})^{\ell-1} \le \exp\left(-(\ell-1)\exp\left(-2Ce/\epsilon\right)\right).$$

The above completes the proof of (55). Then we proceed as follows:

$$\begin{aligned}
\mathbb{E}\left[|X_T \oplus Y_T| \cdot \mathbf{1}\{\mathcal{E}_T\}\right] \le \mathbb{E}\left[H_{\le T}\mathbf{1}\{\mathcal{E}_T\}\right] &\le \sum_{\ell=\Delta^{2/3}}^{n} \ell \cdot \mathbf{Pr}\left[H_{\le T} = \ell\right] \\
&\le \Delta^{2/3} \cdot \mathbf{Pr}\left[H_{\le T} \ge \Delta^{2/3}\right] + \sum_{\ell=\Delta^{2/3}+1}^{n} \mathbf{Pr}\left[H_{\le T} \ge \ell\right] \\
&< \Delta^{2/3} \sum_{\ell=\Delta^{2/3}}^{n} \mathbf{Pr}\left[H_{\le T} \ge \ell\right] \\
&< \Delta^{2/3} \sum_{\ell=\Delta^{2/3}}^{n} \exp\left(-(\ell-1)\exp\left(-6C/\epsilon\right)\right) \qquad \text{(from (55))} \\
&\le 2\Delta^{2/3}\exp(-\Delta^{2/3}e^{-6C/\epsilon}). \qquad (56)
\end{aligned}$$

Note that the above quantity is at most $\exp(-\sqrt{\Delta})$ for large Δ. This completes the proof. ∎

Proof of Lemma 23.4. For this proof we use Lemma 21. We consider the contribution to the expectation $\mathbb{E}\left[|S_{T\log\Delta}|\right]$ from the vertices inside the ball $B_R(v^*)$ and the vertices outside the ball, i.e., $V \setminus B_R(v^*)$, where $R = \sqrt{\Delta}$.

First consider the vertices in $B_R(v^*)$. Lemma 21 implies that some vertex $w \in B_R(v^*)$ at time $T' = T\log\Delta \le \exp(\Delta/C)$ is $(50, 2\Delta^{3/5})$-nice with probability at least $1 - \exp(-\Delta/C)$. This observation implies that

$$\mathbb{E}\left[|S_{T\log\Delta} \cap B_R(v^*)|\right] \le \exp(-\Delta/C)|B_R(v^*)| \le \exp\left(-4\sqrt{\Delta}\right). \qquad (57)$$

To bound the number of disagreements outside $B_R(v^*)$, we observe that each such disagreement comes from a path of disagreements which starts from v^*. Such a path of disagreements is of length at least R. This observation implies that $\mathbb{E}[|S_{T\log\Delta} \cap \bar{B}_R(v^*)|]$ is upper bounded by the expected number of disagreements that start from v^* and have length at least R.

Note that there are at most Δ^ℓ many paths of disagreement of length ℓ that start from v^*. Furthermore, for a fixed path of length ℓ to become path of disagreement up to time $T\log\Delta$, there should be ℓ updates which turn its vertices into disagreeing. Each vertex is chosen to be updated with probability $1/n$, while it becomes disagreeing with probability at most e/Δ. Hence we have the following:

$$\begin{aligned}
\mathbb{E}\left[|S_{T\log\Delta}\cap\bar{B}_R(v^*)|\right] &\le \sum_{\ell\ge R}\Delta^\ell\binom{T\log\Delta}{\ell}\left(\frac{e}{n\Delta}\right)^\ell \\
&\le \sum_{\ell\ge R}\left(\frac{e^2T\log\Delta}{\ell n}\right)^\ell \qquad\qquad \text{(since } \tbinom{n}{s}\le (ne/s)^s\text{)} \\
&\le \sum_{\ell\ge R}\left(\frac{e^2C\log\Delta}{\ell\epsilon}\right)^\ell \\
&\le (1/20)^{\sqrt{\Delta}} \le \exp\left(-2\sqrt{\Delta}\right). \qquad (58)
\end{aligned}$$

Summing the bound of $\mathbb{E}\left[|S_{T\log\Delta}\cap B_R(v^*)|\right]$ and $\mathbb{E}\left[|S_{T\log\Delta}\cap\bar{B}_R(v^*)|\right]$ from (57) and (58), respectively, gives the desired bound for $\mathbb{E}\left[|S_{T\log\Delta}|\right]$. ∎

5.5. Proof of Theorem 19. Fix v and R as specified in the statement of the theorem. Recall that for X_t, Y_t we let $D_t = \{w : X_t\oplus Y_t\}$ and denote $H(X_t,Y_t) = |D_t|$. That is, $H(X_t,Y_t)$ is the Hamming distance between X_t, Y_t. We let the accumulative difference be

$$D_{\le t} = \bigcup_{t'\le t} D_t.$$

Also, let $H_{\le t} = |D_{\le t}|$. Recall that we define the distance between the two chains X_t, Y_t as follows:

$$\mathcal{D}(X_t,Y_t) = \sum_{u\in X_t\oplus Y_t}\Phi(u),$$

where $\Phi : V\to[1,12]$ is defined in Theorem 9. The metric $\mathcal{D}(X_t,Y_t)$ generalizes the Hamming metric in the following sense: the disagreement in each vertex v instead of contributing one it contributes $\Phi(v)$. Since $\Phi(u)\ge 1$, for every $u\in V$, for any two X_t, Y_t we always have

$$\mathcal{D}(X_t,Y_t)\ge H(X_t,Y_t). \qquad (59)$$

For proving the theorem we use the following result which relates the uniformity property with convergence w.r.t. the metric $\mathcal{D}$ we define above.

LEMMA 24. *For $\delta>0$, let sufficiently small $\epsilon=\epsilon(\delta)$ and $\Delta\ge\Delta_0$. Consider a graph $G=(V,E)$ of maximum degree Δ and let $\lambda\le(1-\delta)\lambda_c(\Delta)$. Also, let $(X_t),(Y_t)$ be the Glauber dynamics on the hard-core model with fugacity λ and underlying graphs G.*

For some time t, assume that $X_t\oplus Y_t=\{v^*\}$ for some $v^*\in V$ such that

$$\mathbf{W}_{X_t}(v^*)\le\sum_{z\in N(v^*)}\omega^*(z)\cdot\Phi(z)+\epsilon\Delta; \qquad (60)$$

$\mathbf{W}_{X_t}(v)$ is defined in (18). Then, coupling the chains maximally we have that

$$\mathbb{E}\left[\mathcal{D}(X_{t+1},Y_{t+1})-\mathcal{D}(X_t,Y_t)\mid X_t,Y_t\right] < -c/n$$

for appropriate $c=c(\epsilon,\delta)>0$.

Proof. Let $\Phi_{\max}=\max_{z\in V}\Phi(z)$, where $\Phi: V(G)\to\mathbb{R}_{\ge 0}$, as in Theorem 9. Each vertex $v\in V$ is called a "low degree vertex" if $\deg(v)\le\hat{\Delta}=\frac{\Delta}{2e\cdot\Phi_{\max}}$.

For a low degree vertex v^* it turns out that assumption (60) is not particularly useful. This follows from the observation that the quantity $\epsilon\Delta$ may be greater than the actual degree of the vertex v^*. Then, the information we get from (60) about the number of unblocked neighbors of v^* becomes trivial. However, the assumption that the degree of v^* is small, by itself, is sufficient to yield the desirable result.

It holds that

$$\mathbb{E}\left[\mathcal{D}(X_{t+1}, Y_{t+1}) - \mathcal{D}(X_t, Y_t)\right] \leq -\frac{\Phi(v^*)}{n} + \frac{1}{n}\sum_{z\in N(v^*)}\frac{\lambda}{1+\lambda}\Phi(z).$$

We get the inequality above by working as follows: The distance between the two chains changes when we updated either v^* or some vertex $z \in N(v^*)$.

With probability $1/n$ the the update involves the vertex v^*. Since there is no disagreement at the neighborhood of v^* we can couple X_t and Y_t such that $X_{t+1}(v^*) = Y_{t+1}(v^*)$ with probability 1. That is, the distance between the chain decreases by $\Phi(v^*)$.

We make the (worst case) assumption that all the vertices in $N(v^*)$ are unblocked and unoccupied. We have a new disagreement between the two chains, i.e., an increase in the distance, only if some vertex $z \in N(v^*)$ is chosen to be updated and one of the chains sets z occupied. Since $X_t(v^*) \neq Y_t(v^*)$ one of the chains cannot set z occupied. Each $z \in N(v^*)$ is chosen with probability $1/n$ and it is set occupied by one the two chains with probability $\frac{\lambda}{1+\lambda}$. Then, the distance between the chains increases by $\Phi(z)$. Then we get the following:

$$\begin{aligned}\mathbb{E}\left[\mathcal{D}(X_{t+1}, Y_{t+1}) - \mathcal{D}(X_t, Y_t)\right] &\leq -\frac{\Phi(v^*)}{n} + \frac{1}{n}\sum_{z\in N(v^*)}\frac{\lambda}{1+\lambda}\Phi(z)\\ &\leq -\frac{1}{n}\left(\Phi(v^*) - \Phi_{\max}\cdot(1-\delta)\lambda_c(\Delta)\cdot\hat{\Delta}\right)\\ &\leq -\frac{1}{n}\left(\Phi(v^*) - 1/2\right) \leq -1/(2n),\end{aligned} \tag{61}$$

where the last inequality follows from the fact that $1 \leq \Phi(u) \leq 12$ for every $u \in V$, $\hat{\Delta} = \frac{\Delta}{2e\cdot\Phi_{\max}}$, and $\lambda \leq e/\Delta$. For the case where v is a high degree vertex we have the following:

$$\mathbb{E}\left[\mathcal{D}(X_{t+1}, Y_{t+1}) - \mathcal{D}(X_t, Y_t)\right] \leq -\frac{\Phi(v^*)}{n} + \frac{1}{n}\sum_{z\in N(v^*)}\frac{\lambda}{1+\lambda}\omega^*(z)\Phi(z) + \frac{1}{n}\frac{\lambda}{1+\lambda}\epsilon\Delta.$$

We get the inequality above by working as follows. As before, the interesting cases are those where the update involves the vertex v^* or $N(v^*)$. As we argued above when the vertex v^* is updated the distance between the two chains decreases by $\Phi(v^*)$.

As far as the neighbors of v^* are concerned we observe the following: If some $z \in N(v^*)$ is blocked, then with probability 1 is set unoccupied in both chains. This means that $X_{t+1}(z) = Y_{t+1}(z)$, i.e., the distance between the two chains remains unchanged. If the update involves an unblocked vertex $z \in N(v^*)$, then with probability $\frac{\lambda}{1+\lambda}$ the vertex z becomes occupied at only one of the two chains and the distance between the chains increases by $\Phi(z)$. The assumption (60) implies that the expected contribution from the unblocked neighbors of v^* is

$$\frac{1}{n}\frac{\lambda}{1+\lambda}\mathbf{W}_{X_t}(v^*) \leq \frac{1}{n}\sum_{z\in N(v^*)}\frac{\lambda}{1+\lambda}\omega^*(z)\Phi(z) + \frac{1}{n}\frac{\lambda}{1+\lambda}\epsilon\Delta.$$

Then we get that

$$
\begin{aligned}
&\mathbb{E}\left[\mathcal{D}(X_{t+1},Y_{t+1})-\mathcal{D}(X_t,Y_t)\mid X_t,Y_t\right]\\
&\quad\le -\frac{\Phi(v^*)}{n}+\frac{1}{n}\sum_{z\in N(v^*)}\frac{\lambda}{1+\lambda}\omega^*(z)\Phi(z)+\frac{1}{n}\frac{\lambda}{1+\lambda}\epsilon\Delta\\
&\quad\le -\frac{1}{n}\left(\Phi(v^*)-\sum_{z\in N(v^*)}\frac{\lambda}{1+\lambda}\omega^*(z)\Phi(z)-e\epsilon\right)\\
&\quad\le -\frac{1}{n}\left(\Phi(v^*)-\sum_{z\in N(v^*)}\frac{\lambda}{1+\omega^*(z)\lambda}\omega^*(z)\Phi(z)-e\epsilon-\lambda^2\right) \qquad (62)\\
&\quad\le -\frac{1}{n}\left(\delta\Phi(v^*)/6-e\epsilon-\lambda^2\right) \qquad \text{(by (13))}\\
&\quad\le -c/n, \qquad (63)
\end{aligned}
$$

where (62) follows since $\omega^*(z)\in[0,1]$ and $1\le\Phi(z)\le 12$. The last inequality follows by using that $\lambda<e/\Delta$ and by taking sufficiently small $\epsilon>0$ and large Δ.

The lemma follows from (61) and (63). ∎

We start by proving statement 1 of Theorem 19.

Proof of Theorem 19.1. In this proof we use the uniformity result stated in Theorem 18. Let

$$T_b=\max\{C_b n, C_a n\},$$

where the quantities C_b, C_a are from Lemma 22 and Theorem 18, respectively.

Since $T_m\le n\exp(\Delta/C)$, we can apply Theorem 18 to conclude that the desired local uniformity properties holds with high probability for all $t\in I:=[T_b,T_m]$.

For $t\ge T_b$ we define the following *bad* events:

- $\mathcal{E}(t)$ denotes the event that at some time $s<t$, it holds that $H_s>\Delta^{2/3}$, where $H_s=|X_s\oplus Y_s|$.
- $\mathcal{B}_1(t)$ denotes the event that $D_{\le t}\not\subseteq B_{\sqrt{\Delta}}(v^*)$.
- $\mathcal{B}_2(t)$ denotes the event that there exists a time $T_b\le\tau\le t$, $z\in B_{\sqrt{\Delta}}(v^*)$ such that

$$\mathbf{W}_{X_t}(z)>\Theta(z,\epsilon)=\sum_{u\in N(z)}\omega^*(u)\Phi(u)+\epsilon\Delta,$$

where $\omega^*\in[0,1]^V$ is defined in Lemma 4 and $\Phi:V\to[1,12]$ is defined in Theorem 9.

Also, we let the event

$$\mathcal{B}(t)=\mathcal{B}_1(t)\cup\mathcal{B}_2(t),$$

while we let the "good" event

$$\mathcal{G}(t)=\bar{\mathcal{E}}(t)\cap\bar{\mathcal{B}}(t).$$

We follow the convention that we drop the time t for all of the above events when we are referring to the event at time T_m.

We bound the Hamming distance by conditioning on the above event in the following manner:

$$\begin{aligned}\mathbb{E}\left[H_{T_m}\right] &= \mathbb{E}\left[H_{T_m}\mathbf{1}\{\mathcal{E}\}\right] + \mathbb{E}\left[H_{T_m}\mathbf{1}\{\bar{\mathcal{E}}\}\mathbf{1}\{\mathcal{B}\}\right] + \mathbb{E}\left[H_{T_m}\mathbf{1}\{\bar{\mathcal{E}}\}\mathbf{1}\{\bar{\mathcal{B}}\}\right]\\ &\le \mathbb{E}\left[H_{T_m}\mathbf{1}\{\mathcal{E}\}\right] + \Delta^{2/3}\mathbf{Pr}\left[\mathcal{B}\right] + \mathbb{E}\left[H_{T_m}\mathbf{1}\{\mathcal{G}\}\right]\\ &\le \exp(-\sqrt{\Delta}) + \Delta^{2/3}\mathbf{Pr}\left[\mathcal{B}\right] + \mathbb{E}\left[H_{T_m}\mathbf{1}\{\mathcal{G}\}\right], \end{aligned} \tag{64}$$

where in the last inequality we used Lemma 23.3. For the second term in the (64) we prove the following:

$$\mathbf{Pr}\left[\mathcal{B}\right] \le \exp\left(-\sqrt{\Delta}\right). \tag{65}$$

Finally, for the third term in the (64) we prove the following:

$$\mathbb{E}\left[H_{T_m}\mathbf{1}\{\mathcal{G}\}\right] \le 1/9. \tag{66}$$

Part 1 of the theorem follows by plugging into (64) the bounds in (65) and (66). ∎

Proof of (65). We can bound the probability of the event $\mathcal{B}_1$ by a standard paths of disagreement argument. We are looking at the probability of a path of disagreement of length $\ell = \sqrt{\Delta}$, within $T_m = C'n/\epsilon$ steps, and hence

$$\begin{aligned}\mathbf{Pr}\left[\mathcal{B}_1\right] &\le \Delta^{\ell}\binom{T_m}{\ell}\left(\frac{e}{n\Delta}\right)^{\ell}\\ &\le \left(e^2C'/\epsilon\right)^{\ell} \qquad\qquad (\text{since } \tbinom{N}{i} \le (Ne/i)^i)\\ &\le \exp\left(-2\sqrt{\Delta}\right). \end{aligned} \tag{67}$$

We can bound the probability of the event $\mathcal{B}_2$ by working as follows: The assumption is that v is $(200, R)$-nice for radius $R \ge \Delta^{3/5}$. Then, each vertex $z \in B_{\sqrt{\Delta}}(v^*)$ is $(400, R')$-nice for the constant radius $R'(\gamma, \delta)$ required for the statement for the hypothesis of Theorem 18. Therefore, in the interval $I = [T_b, T_m]$ the uniformity condition for each vertex z fails with probability at most $\exp(-\Delta/C)$. More precisely, we have that

$$\mathbf{Pr}\left[\mathcal{B}_2\right] \le \exp(-\Delta/C)\Delta^{\sqrt{\Delta}+1} \le \exp\left(-2\sqrt{\Delta}\right). \tag{68}$$

Using a simple union bound, we get that $\mathbf{Pr}\left[\mathcal{B}\right] \le \mathbf{Pr}\left[\mathcal{B}_1\right] + \mathbf{Pr}\left[\mathcal{B}_2\right]$. Then (65) follows by plugging (67) and (68) into the union bound. ∎

Proof of (66). Recall that for the two chains X_t, Y_t we defined the following notion of distance:

$$\mathcal{D}(X_t, Y_t) = \sum_{w \in X_t \oplus Y_t} \Phi(w).$$

Note that for every $z \in V$ it holds that $1 \le \Phi(z) \le 12$. This implies that we always have that $\mathcal{D}(X_t, Y_t) \ge H(X_t, Y_t)$, where $H(X_t, Y_t)$ is the Hamming distance between X_t, Y_t. For showing that (66) indeed holds, it suffices to show that

$$\mathbb{E}\left[\mathcal{D}(X_{T_m}, Y_{T_m})\mathbf{1}\{\mathcal{G}\}\right] \le 1/9. \tag{69}$$

Let $Q_0 = X_t, Q_1, Q_2, \ldots, Q_h = Y_t$ be a sequence of independent sets where $h = |X_t \oplus Y_t|$ and Q_{i+1} is obtained from Q_i by changing the assignment of one vertex w_i

from $X_t(w_i)$ to $Y_t(w_i)$. We maximally couple Q_i and Q_{i+1} in one step of the Glauber dynamics to obtain Q'_i and Q'_{i+1}. More precisely, both chains update the spin of the same vertex and maximize the probability of choosing the same new assignment for the chosen vertex.

Consider a pair Q_i, Q_{i+1}. Note that Q_i, Q_{i+1} differ only on the assignment of w_i. With probability $1/n$ both chains update the spin of vertex w_i. Since all of the neighbors of w_i have the same spin, with probability 1 we assign the same spin on w_i in both chains. Such an update reduces the distance of the two chains by $\Phi(w_i)$.

Consider now the update of vertex $w \in N(w_i)$. Also, without loss of generality assume that $Q_i(w_i)$ is occupied while $Q_{i+1}(w_i)$ is unoccupied. It is direct that the worst case is when w is unblocked in the chain Q_{i+1}. Otherwise, i.e., if w is blocked, then with probability 1 we have $Q'_{i+1}(w) = Q'_i(w) =$"unoccupied."

Assuming that w_i is blocked in the chain Q_i and unblocked in the chain Q_{i+1}, we get $Q'_i(w) \neq Q'_{i+1}(w)$ if the coupling chooses to set w_i occupied in Q'_{i+1}. Otherwise, we have $Q'_i(w) = Q'_{i+1}(w)$. Therefore, the disagreement happens with probability $\leq \frac{\lambda}{1+\lambda} < e/\Delta$, where the last inequality holds for $\lambda < \lambda_c$ and $\Delta \geq \Delta_0$.

Therefore, given Q_i, Q_{i+1}, we have that

$$\mathbb{E}\left[\mathcal{D}(Q'_i, Q'_{i+1}) - \mathcal{D}(Q_i, Q_{i+1})\right] \leq -\frac{\Phi(w_i)}{n} + \frac{e}{n\Delta} \sum_{z \in N(w_i)} \Phi(z). \tag{70}$$

Since we have that $1 \leq \Phi(z) \leq 12$, for any $z \in V$ and $|N(v)| \leq \Delta$, we get the trivial bound that

$$\mathbb{E}\left[\mathcal{D}(Q'_i, Q'_{i+1}) - \mathcal{D}(Q_i, Q_{i+1})\right] \leq 35/n.$$

Therefore,

$$\mathbb{E}\left[\mathcal{D}(X_{t+1}, Y_{t+1})\right] \leq (1 + 35/n)\, \mathcal{D}(X_t, Y_t). \tag{71}$$

The above bound is going to be used only for the burn-in phase, i.e., the first T_b steps. We use a significantly better bound for the remaining $T_m - T_b$ steps.

Since the event $\mathcal{G}$ holds, for all $0 \leq i \leq h$, $z \in B_R(v^*)$ and all $t \in [T_b, T_m - 1]$, we have that

$$\mathbf{W}_{Q_i}(z) \leq \Theta(z, \epsilon) + \Delta^{2/3} \leq \Theta(z, 2\epsilon). \tag{72}$$

The first inequality follows from our assumption that both event $\bar{\mathcal{E}}$ and $\bar{\mathcal{B}}_2$ occur. The second follows from the definition of the quantity Θ.

Using Lemma 24 we get the following: For Q_i, Q_{i+1} which satisfy (72) it holds that

$$\mathbb{E}\left[\mathcal{D}(Q'_i, Q'_{i+1})\right] \leq (1 - C'/n)\, \mathcal{D}(Q_i, Q_{i+1})$$

for appropriately chosen C'. The above inequality implies the following: given X_t, Y_t and assuming that $\mathcal{G}(t)$ holds, we get that

$$\mathbb{E}\left[\mathcal{D}(X_{t+1}, Y_{t+1})\right] \leq (1 - C/n)\, \mathcal{D}(X_t, Y_t). \tag{73}$$

Let $t \in [T_b, T_m - 1]$. Then we have that

$$\begin{aligned}
\mathbb{E}\left[\mathcal{D}(X_{t+1}, Y_{t+1})\mathbf{1}\{\mathcal{G}(t)\}\right] &= \mathbb{E}\left[\mathbb{E}\left[\mathcal{D}(X_{t+1}, Y_{t+1})\mathbf{1}\{\mathcal{G}(t)\} \mid X_0, Y_0, \ldots, X_t, Y_t\right]\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[\mathcal{D}(X_{t+1}, Y_{t+1}) \mid X_0, Y_0, \ldots, X_t, Y_t\right]\mathbf{1}\{\mathcal{G}(t)\}\right] \\
&\leq (1 - C/n)\, \mathbb{E}\left[\mathcal{D}(X_t, Y_t)\mathbf{1}\{\mathcal{G}(t)\}\right] \\
&\leq (1 - C/n)\, \mathbb{E}\left[\mathcal{D}(X_t, Y_t)\mathbf{1}\{\mathcal{G}(t-1)\}\right].
\end{aligned}$$

The first equality is Fubini's theorem, and the second equality is due to the fact that $X_0, Y_0, \dots X_t, Y_t$ determine uniquely $\mathcal{G}(t)$. The first inequality uses (73) while the second inequality uses the fact that $\mathcal{G}(t) \subset \mathcal{G}(t-1)$. By induction, it follows that

$$\mathbb{E}\left[\mathcal{D}(X_{T_m}, Y_{T_m})\mathbf{1}\{\mathcal{G}(T_m)\}\right] \leq (1 - C/n)^{T_m - T_b}\,\mathbb{E}\left[\mathcal{D}(X_{T_b}, Y_{T_b})\mathbf{1}\{\mathcal{G}(T_b)\}\right].$$

Using the same arguments and (71) for $\mathbb{E}\left[\mathcal{D}(X_{T_b}, Y_{T_b})\mathbf{1}\{\mathcal{G}(T_b)\}\right]$ we get that

$$\mathbb{E}\left[\mathcal{D}(X_{T_m}, Y_{T_m})\mathbf{1}\{\mathcal{G}(T_m)\}\right] \leq (1 - C/n)^{T_m - T_b}\,(1 + 35/n)^{T_b}\,\mathcal{D}(X_0, Y_0). \tag{74}$$

The result follows from the choice of constants and noting that $\mathcal{D}(X_0, Y_0) < 12$. ∎

Proof of Theorem 19.2. Recall from the proof of Theorem 19.1 that $\mathcal{B}_1$ is the event that $D_{\leq T_m} \not\subseteq B_{\sqrt{\Delta}}(v^*)$. Also consider $\mathcal{J}_1$ to be the event that $D_{T_m} \not\subseteq B_{\sqrt{\Delta}}(v^*)$. Noting that $\mathcal{J}_1 \subset \mathcal{B}_1$, we get that

$$\mathbf{Pr}\left[\mathcal{J}_1\right] \leq \mathbf{Pr}\left[\mathcal{B}_1\right] \leq \exp\left(-\sqrt{\Delta}\right),$$

where the last inequality follows from (67).

Let $\mathcal{J}_2$ be the event that X_{T_m} or Y_{T_m} has a vertex $w \in B_{\sqrt{\Delta}}(v^*)$ which is not $(50, r)$-nice, where $r = R - \sqrt{\Delta} - 2$. By the hypothesis of Theorem 19, each vertex $w \in B_{\sqrt{\Delta}}(v^*)$ is $(400, r)$-nice for radius $r = R - \sqrt{\Delta}$ in both X_0 and Y_0. Therefore, by Lemma 22, each vertex $w \in B_{\sqrt{\Delta}}(v^*)$ is $(50, r)$-nice for radius $r = R - \sqrt{\Delta} - 2$ in X_{T_m} and Y_{T_m} with probability $\geq 1 - \exp(-\Delta/(C_b \log \Delta))$. Therefore, by a union bound over the vertices in $B_{\sqrt{\Delta}}(v^*)$ we have that

$$\Pr[\mathcal{J}_2] \leq \exp(-\Delta/(2C_b \log \Delta)).$$

Theorem 19.2. follows by noting that $\Pr[\mathcal{Z}] \leq \Pr[\mathcal{J}_1] + \Pr[\mathcal{J}_2]$. ∎

5.6. Proof of Lemma 17.

Proof of Lemma 17. The proof of the lemma is identical to the proof of [4, Lemma 12]. We present it here to illustrate how the various results in this paper combine together.

For proving the lemma, the basic idea is to combine Lemma 23, Theorem 19, and path coupling. To be more specific, we partition the time interval $[0, T]$ into epochs, and each one is of length $T_m = C'n/\epsilon$. For each epoch $i \geq 1$ we analyze the expected number of disagreements, i.e., $\mathbb{E}[|X_{iT_m} \oplus Y_{iT_m}|]$. Given X_{iT_m}, Y_{iT_m} we analyze $\mathbb{E}[|X_{(i+1)T_m} \oplus Y_{(i+1)T_m}|]$ by using path coupling. Path coupling considers a sequence of configurations $Z_0, \dots, Z_\ell$ for some ℓ, such that $Z_0 = X_{iT_m}$, $Z_\ell = Y_{iT_m}$ and each Z_j, Z_{j+1} differ on the assignment of a single vertex w_j. We couple each Z_j, Z_{j+1} for T_m steps and we get Z'_j, Z'_{j+1}. Path coupling implies that

$$\mathbb{E}\left[|X_{(i+1)T_m} \oplus Y_{(i+1)T_m}|\right] \leq \sum_{j=0}^{\ell-1} \mathbb{E}\left[|Z'_j \oplus Z'_{j+1}|\right]. \tag{75}$$

We call the sequence $Z_0, \dots, Z_\ell$ as the interpolating sequence. In the rest of the proof when we use the phrase "by path coupling" we imply that we use the interpolating sequence and the above relation to bound expected number of disagreements at a specific moment.

For bounding each $\mathbb{E}[|Z'_j \oplus Z'_{j+1}|]$ we use either Lemma 23 or Theorem 19. The choice depends on whether w_j, the vertex that Z_j, Z_{j+1} disagree, is nice or bad. For the $(i+1)$th epoch, we consider the disagreement between X_{iT_m} and Y_{iT_m} being $(200, R_i)$-bad or not for $R_i = 2\Delta^{3/5} - 2i\sqrt{\Delta}$.

We need to define the following events: Let $\mathcal{E}'_i$ be the event that for some $t \le iT_m$ we have $|X_t \oplus Y_t| \ge \Delta^{2i/3}$. Let $\mathcal{S}_i$ be the event that for some $t \le iT_m$ there exists a $(200, R_i)$-bad disagreement for X_t and Y_t.

Letting $H_{i+1} = |X_{(i+1)T_m} \oplus Y_{(i+1)T_m}|$, we have

$$\mathbb{E}\left[H_{i+1}\right] \le \mathbb{E}\left[H_{i+1}\mathbf{1}\{\mathcal{E}'_i\}\right] + \mathbb{E}\left[H_{i+1}\mathbf{1}\{\bar{\mathcal{S}}_i\}\right] + \mathbb{E}\left[H_{i+1}\mathbf{1}\{\bar{\mathcal{E}}'_i\}\mathbf{1}\{\mathcal{S}_i\}\right]. \tag{76}$$

The lemma follows by bounding appropriately the three summands on the r.h.s. of (76).

As far as $\mathbb{E}\left[H_{i+1}\mathbf{1}\{\mathcal{E}'_i\}\right]$ is concerned, we have that

$$\begin{aligned}
\mathbb{E}\left[H_{i+1}\mathbf{1}\{\mathcal{E}'_i\}\right] &= \mathbb{E}\left[\mathbb{E}\left[H_{i+1}\mathbf{1}\{\mathcal{E}'_i\} \mid X_t, Y_t, \text{ for } t \le iT_m\right]\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[H_{i+1} \mid X_t, Y_t, \text{ for } t \le iT_m\right] \mathbf{1}\{\mathcal{E}'_i\}\right] \\
&\le \exp\left(3C'/\epsilon\right)\mathbb{E}\left[H_i\ \mathbf{1}\{\mathcal{E}'_i\}\right] &(77)\\
&\le \exp\left(3C'/\epsilon\right)\left(\mathbb{E}\left[H_i\ \mathbf{1}\{\mathcal{E}'_{i-1}\}\right] + \mathbb{E}\left[H_i\ \mathbf{1}\{\mathcal{E}'_i\}\mathbf{1}\{\bar{\mathcal{E}}'_{i-1}\}\right]\right), &(78)
\end{aligned}$$

where in (77) we use Lemma 23.1 and path coupling. We proceed by bounding the two terms in the r.h.s. of (78).

Starting with $\mathbb{E}[H_i\ \mathbf{1}\{\mathcal{E}'_i\}\mathbf{1}\{\bar{\mathcal{E}}'_{i-1}\}]$ consider the interpolating sequence $Z_0, \ldots, Z_\ell$ we use for $X_{(i-1)T_m}$ and $Y_{(i-1)T_m}$ so as to bound the expectation of H_i. For every $j = 0, \ldots, \ell-1$, let (Z'_j, Z'_{j+1}) be the pair of configuration after coupling the pair (Z_j, Z_{j+1}) for T_m steps, while $H_{i,j} = |Z'_j \oplus Z'_{j+1}|$. Let $\mathcal{E}'_{i,j}$ be the event that $H_{i,j} \ge \Delta^{2/3}$. For both $\mathcal{E}'_i$ and $\bar{\mathcal{E}}'_{i-1}$ to occur there should be at least one $j \in \{0, \ldots, \ell-1\}$ such that $H_{i,j} \ge \Delta^{2/3}$, i.e., the event $\mathcal{E}'_{i,j}$ occurs.

Noting that $H_i \le \sum_{j=1}^{H_{i-1}} H_{i,j}$ and $\mathcal{E}'_i = \mathcal{E}_{i-1} \cup (\cup_{j=1}^{H_{i-1}} \mathcal{E}'_{i,j})$ we have

$$\begin{aligned}
H_i\mathbf{1}\{\mathcal{E}_i\}\mathbf{1}\{\bar{\mathcal{E}}'_{i-1}\} &\le \sum_{j=1}^{H_{i-1}} H_{i,j} \sum_{k=1}^{H_{i-1}} \mathbf{1}\{\mathcal{E}_{i,k}\}\mathbf{1}\{\bar{\mathcal{E}}'_{i-1}\} \\
&\le \sum_{j,k=1}^{\Delta^{2(i-1)/3}} H_{i,j}\mathbf{1}\{\mathcal{E}_{i,k}\} \\
&\le \Delta^{2(i-1)/3} \sum_{j=1}^{\Delta^{2(i-1)/3}} H_{i,j}\mathbf{1}\{\mathcal{E}_{i,j}\}.
\end{aligned}$$

Applying Lemma 23 for each pair Z_j, Z_{j+1}, path coupling yields

$$\mathbb{E}\left[H_i\ \mathbf{1}\{\mathcal{E}'_i\}\mathbf{1}\{\bar{\mathcal{E}}'_{i-1}\}\right] \le \Delta^{4(i-1)/3}\exp\left(-\sqrt{\Delta}\right). \tag{79}$$

As far as $\mathbb{E}[H_i\ \mathbf{1}\{\mathcal{E}'_{i-1}\}]$ is concerned, we use induction. More specifically, we have

$$\begin{aligned}
\mathbb{E}\left[H_{i+1}\ \mathbf{1}\{\mathcal{E}'_i\}\right] &\le \mathbb{E}\left[H_i\ \mathbf{1}\{\mathcal{E}'_i\}\ \mathbf{1}\{\mathcal{E}'_{i-1}\}\right] + \mathbb{E}\left[H_i\ \mathbf{1}\{\mathcal{E}'_i\}\ \mathbf{1}\{\bar{\mathcal{E}}'_{i-1}\}\right] \\
&\le \mathbb{E}\left[H_i\ \mathbf{1}\{\mathcal{E}'_{i-1}\}\right] + \mathbb{E}\left[H_i\ \mathbf{1}\{\mathcal{E}'_i\}\ \mathbf{1}\{\bar{\mathcal{E}}'_{i-1}\}\right] \\
&\le \mathbb{E}\left[H_i\ \mathbf{1}\{\mathcal{E}'_{i-1}\}\right] + \Delta^{4(i-1)/3}\exp\left(-\sqrt{\Delta}\right). &(80)
\end{aligned}$$

Using the above and noting that $H_0 = 1$ we get

$$\mathbb{E}\left[H_i\ \mathbf{1}\{\mathcal{E}'_{i-1}\}\right] \le \exp\left(3iC'/\epsilon\right)\Delta^{4(i)/3}\exp\left(-\sqrt{\Delta}\right). \tag{81}$$

As far as the second summand in the r.h.s. of (76) we have

$$\begin{aligned}\mathbb{E}\left[H_{i+1}\mathbf{1}\{\bar{\mathcal{S}}_i\}\right] &= \mathbb{E}\left[\mathbb{E}\left[H_{i+1}\mathbf{1}\{\bar{\mathcal{S}}_i\} \mid H_i\right]\right] \\ &\le 3^{-1}\mathbb{E}\left[H_i\mathbf{1}\{\bar{\mathcal{S}}_i\}\right] \\ &\le 3^{-1}\mathbb{E}\left[H_i\mathbf{1}\{\bar{\mathcal{S}}_{i-1}\}\right] \\ &\le 3^{-(i+1)}H_0 = 3^{-(i+1)}. \end{aligned} \tag{82}$$

As far as the third summand in the r.h.s. of (76) we have

$$\begin{aligned}\mathbb{E}\left[H_{i+1}\mathbf{1}\{\bar{\mathcal{E}}'_i\}\mathbf{1}\{\mathcal{S}_i\}\right] &= \mathbb{E}\left[\mathbb{E}\left[H_{i+1} \mid X_{iT_m,Y_{iT_m}}\right]\mathbf{1}\{\bar{\mathcal{E}}'_i\}\mathbf{1}\{\mathcal{S}_i\}\right] \\ &\le \exp\left(3C'/\epsilon\right)\mathbb{E}\left[H_i\mathbf{1}\{\bar{\mathcal{E}}'_i\}\mathbf{1}\{\mathcal{S}_i\}\right] \\ &\le \varDelta^{2i/3}\exp\left(3C'/\epsilon\right)\Pr\left[\mathcal{S}_i \setminus \bar{\mathcal{E}}'_i\right]. \end{aligned} \tag{83}$$

We bound $\Pr[\mathcal{S}_i \setminus \bar{\mathcal{E}}'_i]$ by using Theorem 19.2 to each pair of neighboring independent sets that arise at time jT_m for $j = 0, 1, \ldots, i-1$. Since $\mathcal{E}'_i$ does not occur, there are at most $\varDelta^{2i/3}$ neighboring pairs that we need to consider for each j. For each of these pairs we use Theorem 19.2 to bound the probability that a $(200, R_i)$-bad disagreement is generated within the following T_m steps. Taking a union bound over all of the $\le i\varDelta^{2i/3}$ neighboring pairs we consider we get that

$$\mathbb{E}\left[H_{i+1}\mathbf{1}\{\bar{\mathcal{E}}'_i\}\mathbf{1}\{\mathcal{S}_i\}\right] \le i\exp\left(3C'/\epsilon\right)\varDelta^{4i/3}\exp\left(-\sqrt{\varDelta}\right). \tag{84}$$

Plugging (81), (82), and (84) into (76) we get that

$$\mathbb{E}\left[H_{i+1}\right] \le 3^{-(i+1)} + \exp\left(3C'/\epsilon\right)\varDelta^{5i/3}\exp\left(-\sqrt{\varDelta}\right) \le \left(\sqrt{\varDelta}\right)^{-1}, \tag{85}$$

where the last inequality follows by choosing sufficiently large $\varDelta$. ∎

6. Rapid mixing for random regular (bipartite) graphs. It turns out that the girth restriction of Theorem 1 can be relaxed a bit. The main technical reason why we need girth at least 7 is for establishing Theorem 18, our so-called local uniformity result for the Glauber dynamics. Roughly speaking, local uniformity amounts to showing that the number of unblocked neighbors of a vertex v is concentrated about the quantity $\sum_{z\in N(v)}\omega^*(z)$, where $\omega^* \in [0,1]^V$ are the fixed points of a BP-like system of equations.

The analysis of local uniformity can be carried out for a graph with short cycles, i.e., cycles of length less than 7, as long as these cycles are far apart. Generally, the effect of a short cycle is an increase to the fluctuation of the number of unblocked neighbors of a vertex. If the short cycles in G are far apart from each other, then the cumulative increase in the fluctuation is negligible.

Proof of Theorem 2. For an integer $r > 0$, let $\mathcal{G}_n(\varDelta, r)$ be the family of $\varDelta$-regular graphs on n vertices such that the following holds: For each $G \in \mathcal{G}_n(r)$, any two cycles of length < 7 are at graph distance greater than r from each other.

First, we are going to show the following: there exist $r = r(\delta)$, $\varDelta_0 = \varDelta_0(\delta)$, and $C = C(\delta)$ for all $\varDelta \ge \varDelta_0$, all $\lambda < (1-\delta)\lambda_c(\varDelta)$, all $\epsilon > 0$, and for all $G \in \mathcal{G}_n(\varDelta, r)$ the mixing time of the Glauber dynamics on G satisfies

$$T_{\mathrm{mix}}(\epsilon) \le Cn\log(n/\epsilon).$$

Since we consider regular graphs, the weights we introduce in Theorem 9 are not necessary for the path coupling. Additionally, it is direct to see that once we have established the local uniformity property then the path coupling arguments from section 5.2 hold and imply rapid mixing. Hence, for $G \in \mathcal{G}_n(\mathit{\Delta}, r)$, the only aspect of the rapid mixing proof that changes in the presence of short cycles is proving local uniformity.

For an independent set σ and vertex v, let

$$\mathbf{Q}(\sigma, v) = \sum_{z \in N(v)} \mathbf{U}_{z,v}(\sigma).$$

Uniformity amounts to showing that for appropriate $\gamma > 0$ we have the following: Let (X_t) be the continuous (or discrete) time Glauber dynamics on the hard-core model with fugacity λ. If X_0 is $(400, R)$-nice at $v \in V$ for radius $R = R(\delta, \gamma) > 1$, there is $C_1 > 0$ such that

$$\mathbf{Pr}\left[(\forall t \in \mathcal{I}) \quad \mathbf{Q}(X_t, v) < \sum_{z \in N(v)} \omega^*(z) + \gamma \mathit{\Delta}\right] \geq 1 - \exp\left(-\mathit{\Delta}/C_1\right), \tag{86}$$

where the time interval $\mathcal{I} = [C_1 n, n \exp(\mathit{\Delta}/C_1)]$ and ω^* is defined in Lemma 4.

Let $\mathcal{Y}$ contain the set of vertices in G which do not belong to any short cycle, namely, any cycle of length < 7. For each vertex u and some independent set σ, we define

$$\mathbf{Z}(\sigma, u) = \prod_{z \in \hat{N}(u)} \mathbf{Pr}\left[z \notin Y \mid u \notin Y,\ Y(S_2(u)) = \sigma(S_2(u))\right],$$

where $\hat{N}(u) \subseteq N(u)$ contains every $w \in N(u)$ such that $w \in \mathcal{Y}$ and $N(w) \subset \mathcal{Y}$. Recall from (21) the quantity $\mathbf{R}(\sigma, u)$. The difference between the quantities $\mathbf{R}(\sigma, u)$ and $\mathbf{Z}(\sigma, u)$ is that the former considers $N(u)$ and the later considers $\hat{N}(u)$. Note that $|N(u) \setminus \hat{N}(u)| \leq 2$.

To get some intuition why we choose to define $\mathbf{Z}(\sigma, u)$ consider the following. Let $\mathit{\Lambda}$ be the set of vertices that are reachable from u through paths of length 2 that don't use vertices in $N(u) \setminus \hat{N}(u)$. Then, the subgraph that is induced by $\mathit{\Lambda}$ is a tree. The local tree like neighborhood that we used to establish uniformity in Theorem 18 is now replaced by $\mathit{\Lambda}$.

To establish (86) we work similarly to the proof of Theorem 18. That is, first we show the following: Let (X_t) be the continuous time Glauber dynamics on the hard-core model with fugacity λ and underlying graph G. Then there exists $C_1 > 0$, such that for any X_0 which is $(400, R)$-nice at $v \in V$ we have that

$$\mathbf{Pr}\left[(\forall t \in \mathcal{I}) \quad |\mathbf{Z}(X_t, v) - \omega^*(v)| \leq \gamma/10\right] \geq 1 - \exp\left(-20\mathit{\Delta}/C_1\right). \tag{87}$$

To obtain (87) we use the following: Let (X_t) be the continuous time Glauber dynamics on the hard-core model. Assume that X_0 is $(400, R')$-nice at $w \in V$ for radius $R' \leq \mathit{\Delta}^{9/10}$. Then, for $x \in B_{R/2}(v)$ and $I = [t_0, t_1]$, where $t_0 = Cn$, there exists $\hat{C} > 0$ such that

$$\begin{aligned} &\mathbf{Pr}\left[(\forall t \in I) \quad \left|\mathbf{Z}(X_t, x) - \exp\left(-\frac{\lambda}{1+\lambda} \sum_{z \in \hat{N}(x)} \mathbb{E}_{t_z}\left[\mathbf{Z}(X_{t_z}, z)\right]\right)\right| \leq \gamma^2 \delta/40\right] \\ &\qquad \geq 1 - \left(1 + \frac{t_1 - t_0}{n}\right) \exp\left(-\mathit{\Delta}/\hat{C}\right), \end{aligned} \tag{88}$$

where $\mathbb{E}_{t_z}[\mathbf{Z}(X_{t_z}, z)]$ is the expectation w.r.t. random time t_z, the last time that vertex z is updated prior to time t. Equation (88) follows by using arguments very similar to those we used in the proof of Lemma 33.

In light of (88), (87) follows by working as in the proof of Lemma 32. Let us be more specific. Consider some time interval $\mathcal{I}'$ which starts prior to $\mathcal{I}$. Assume that for every $t \in \mathcal{I}'$, (88) holds for every $z \in B_R(v)$. A union bound implies that this holds with probability at least $1 - \exp(-2\Delta/\hat{C})$.

Furthermore, consider some integer $i < R$. Assume that there exists some $s \in \mathcal{I}' \setminus \mathcal{I}$ such for any $t \geq s$, for every vertex $z \in B_{i+1}(v) \cap \mathcal{Y}$ we have that

$$|\Psi\left(\mathbb{E}_{t_z}\left[\mathbf{Z}(X_{t_z}, z)\right]\right) - \Psi(\omega^*(z))| \leq \beta$$

for some $\beta > \epsilon/20$. Then, in the heart of the proof of (87) we have the following contraction property: For every $u \in B_i(v)$, it holds that

$$|\Psi(\mathbf{Z}(X_t, u)) - \Psi(\omega^*(u))| \leq (1 - \delta/24)\beta. \tag{89}$$

To show (89), we note that $\lambda \leq e/\Delta$ for $\Delta \geq \Delta_0$, and for any vertex w we have $N(w) \setminus \hat{N}(w) \leq 2$ and get that

$$|\Psi(\mathbf{Z}(X_t, u)) - \Psi(\omega^*(u))| \leq (1/10)\gamma^2\delta + \left|\Psi\left(\prod_{z \in N(u)}\left(1 - \frac{\lambda}{1+\lambda}\tilde{\omega}(z)\right)\right) - \Psi(\omega^*(u))\right|,$$

where $\tilde{\omega}$ is such that for every $z \in \mathcal{Y}$ we have $\tilde{\omega}(z) = \mathbb{E}_{t_z}[\mathbf{Z}(X_{t_z}, z)]$ while for every $z \notin \mathcal{Y}$ we have $\tilde{\omega}(z) = \omega^*(z)$. The actual derivations for getting the above are very similar to those from (118) until (121) in the proof of Lemma 32. Furthermore, the above inequality implies that

$$\begin{aligned}|\Psi(\mathbf{Z}(X_t, w)) - \Psi(\omega(z))| &\leq (1/10)\gamma^2\delta + D_{v,i}(F(\tilde{\omega}), \omega^*)\\ &\leq (1/10)\gamma^2\delta + (1 - \delta/6)\beta \leq (1 - \delta/24)\beta,\end{aligned}$$

where $F(\cdot)$ is the function defined in (3). The second derivation follows from Lemma 5, while the last follows by having small $\gamma > 0$. Therefore (89) follows. Given the above contraction, we get (87) by following very similar arguments to those we used for Lemma 32.

Additionally to (87), we need to define

$$\mathbf{B}(X_t, v) = \sum_{z \in \hat{N}(v)} \mathbf{U}_{z,v}(X_t).$$

As opposed to $\mathbf{Q}(X_t, v)$, the quantity $\mathbf{B}(X_t, v)$ considers only the neighbors of v which belong to $\hat{N}(v)$.

Using arguments identical to those in the proof of Lemma 34 we get the following: Let (X_t) be the continuous time Glauber dynamics on the hard-core model with fugacity λ. Assume that X_0 is $(400, R)$-nice at v. Then, there is $\hat{C}_1 = \hat{C}_1(\epsilon) > 0$ such that for any $t \in \mathcal{I}$ we have

$$\mathbf{Pr}\left[\left|\mathbf{B}(X_t, v) - \sum_{z \in \hat{N}(v)} \mathbb{E}_{t_z}\left[\mathbf{Z}(X_{t_z}, z)\right]\right| > (\gamma/20)\Delta\right] < \exp\left(-10\Delta/\hat{C}_1\right).$$

Combining the above with (87) in the same way as in the proof of Theorem 18, we get the following: Let (X_t) be the continuous (or discrete) time Glauber dynamics on the hard-core model. If X_0 is $(400, R)$-nice at $v \in V$, there exists $C_1 = C_1(\delta, \epsilon) > 0$ such that

$$\mathbf{Pr}\left[(\forall t \in \mathcal{I}) \quad \mathbf{B}_{X_t}(v) < \sum_{z \in \hat{N}(v)} \omega^*(z) + (\gamma/2)\Delta\right] \geq 1 - \exp\left(-\Delta/C_1\right). \tag{90}$$

Then, we get (90) by noting that it always holds that $|\mathbf{B}(X_t, v) - \mathbf{Q}(X_t, v)| \leq 2\max_z\{\omega^*(z)\} \leq 2$, since $\omega^*(z) \leq 1$.

With all of the above, we conclude that indeed we have rapid mixing for every $G \in \mathcal{G}_n(\Delta, r)$. In light of this conclusion, we prove the theorem by showing that the typical instances of the random graphs we consider in our theorem belong to $\mathcal{G}_n(\Delta, r)$ for any fixed integer $r > 0$. Since the arguments we employ for random regular graphs and random regular bipartite graphs are very similar with each other, our focus will be on random regular graphs.

Let G be a random regular graph of degree Δ. For each integer $r > 0$, let S_r be the family of subsets of vertices of G with cardinality at most r. Furthermore, let $S'_r \subseteq S_r$ contain each $\Lambda \in S_r$ such that the vertices of Λ span a number of edges which is greater than the cardinality of Λ.

Note that if there is a pair of cycles in G of length ℓ_1, ℓ_2 which are at distance ℓ_3 from each other, then $S'_{\ell_1+\ell_2+\ell_3} \neq \emptyset$.

Let Y_r be the cardinality of the set S'_r in G. Following some standard but tedious derivations (e.g., see [17, section 9.2]) we get that $\mathbb{E}[Y_r] = O(n^{-1})$ for any fixed r. Then, applying Markov's inequality we get that with probability $1 - O(n^{-1})$ we have $Y_r = 0$. Clearly, this implies that with probability $1 - O(n^{-1})$ we have that $G \in \mathcal{G}_n(r)$ for any fixed integer $r > 0$.

The theorem follows. ∎

7. BP convergence: Proofs of Propositions 7 and 8. Let $f_{\lambda,d}(x) = (1 + \lambda x)^{-d}$ be the symmetric version of the BP recurrence (3). Let $\hat{x} = \hat{x}(\lambda, d)$ be the unique fixed point of $f_{\lambda,d}(x)$, satisfying $\hat{x}(\lambda, d) = (1 + \lambda\hat{x}(\lambda, d))^{-d}$. We define

$$\alpha(\lambda, d) = \sqrt{\frac{d \cdot \lambda\hat{x}(\lambda, d)}{1 + \lambda\hat{x}(\lambda, d)}}.$$

Proposition 7 in section 2 states that for all $\delta > 0$, there exists $\Delta_0 = \Delta_0(\delta)$, and for all $\Delta \geq \Delta_0$, all $\lambda < (1 - \delta)\lambda_c(\Delta)$, where $\lambda_c(\Delta) = \frac{(\Delta-1)^{\Delta-1}}{(\Delta-2)^{\Delta}}$, it holds that $\alpha(\lambda, \Delta) \leq 1 - \delta/6$.

Proof of Proposition 7. Let $x_0 = \frac{1-\delta/3}{\lambda(\Delta-1+\delta/3)}$. It is easy to verify that

$$\sqrt{\frac{\Delta \cdot \lambda x_0}{1 + \lambda x_0}} \leq 1 - \delta/6.$$

Note that the function $\sqrt{\frac{\Delta\lambda x}{1+\lambda x}}$ is increasing in x. Since $f(x)$ is increasing in λ, it is easy to verify that $\hat{x}(\lambda, d)$ is increasing in λ. We then show that for all $\Delta \geq \Delta_0$, it holds that $\hat{x}(\lambda_0, \Delta) \leq x_0$, where $\lambda_0 = (1 - \delta)\lambda_c(\Delta) = \frac{(1-\delta)(\Delta-1)^{\Delta-1}}{(\Delta-2)^{\Delta}}$, which will prove our proposition.

Since $f_{\lambda_0,\Delta}(x)$ is decreasing in x and $f_{\lambda_0,\Delta}(\hat{x}(\lambda_0, \Delta)) = \hat{x}(\lambda_0, \Delta)$, it is sufficient to show that

$$f_{\lambda_0,\Delta}(x_0) = (1 + \lambda_0 x_0)^{-\Delta} \leq x_0.$$

Note that it holds that

$$\frac{f_{\lambda_0,\Delta}(x_0)}{x_0} = \frac{\lambda_0(\Delta-1+\delta/3)}{(1-\delta/3)(1+\frac{1-\delta/3}{(\Delta-1+\delta/3)})^{\Delta}}$$
$$= \frac{1-\delta}{1-\delta/3}\cdot\frac{(\Delta-1)^{\Delta}(\Delta-1+\delta/3)^{\Delta}}{(\Delta-2)^{\Delta}\Delta^{\Delta}}\cdot\frac{\Delta-1+\delta/3}{\Delta-1}.$$

Therefore, there is a suitable $\Delta_0 = O(\frac{1}{\delta})$ such that for all $\Delta \geq \Delta_0$,

$$\frac{f_{\lambda_0,\Delta}(x_0)}{x_0} \leq \frac{1-\delta}{1-\delta/3}\left(1+O\left(\frac{\eta}{\Delta}\right)\right)\mathrm{e}^{\delta/2.99} < 1,$$

which proves the proposition. ∎

Let $G = (V, E)$ be a graph with maximum degree at most Δ. Assume that $\alpha(\lambda, \Delta) \leq 1$. Recall recurrence F as defined in (3). Proposition 8 in section 2 states that for any $\omega \in [0, 1]^V$, and $v \in V$, it holds that

$$\sqrt{\frac{\lambda F(\omega)(v)}{1+\lambda F(\omega)(v)}}\sum_{u\in N(v)}\sqrt{\frac{\lambda\omega(u)}{1+\lambda\omega(u)}} \leq \alpha(\lambda,\Delta).$$

This proposition was proved implicitly in [20]. We state the proof here in our context for the completeness of the paper.

Proof of Proposition 8. Let $\bar{\omega} \in [0, 1]$ satisfy $1+\lambda\bar{\omega} = (\prod_{u\in N(v)}(1+\lambda\omega(u)))^{\frac{1}{|N(v)|}}$. Denote that $\bar{\nu} = \ln(1 + \lambda\bar{\omega})$ and $\nu(u) = \ln(1 + \lambda\omega(u))$. It then holds that $\bar{\nu} = \frac{1}{|N(v)|}\sum_{u\in N(v)}\nu(u)$. Due to the concavity of $\sqrt{\frac{\mathrm{e}^{\nu}-1}{\mathrm{e}^{\nu}}}$ in ν, by Jensen's inequality,

$$\frac{1}{|N(v)|}\sum_{u\in N(v)}\sqrt{\frac{\lambda\omega(u)}{1+\lambda\omega(u)}} = \frac{1}{|N(v)|}\sum_{u\in N(v)}\sqrt{\frac{\mathrm{e}^{\nu(u)}-1}{\mathrm{e}^{\nu(u)}}} \leq \sqrt{\frac{\mathrm{e}^{\bar{\nu}}-1}{\mathrm{e}^{\bar{\nu}}}} = \sqrt{\frac{\lambda\bar{\omega}}{1+\lambda\bar{\omega}}}.$$

Therefore,

$$\sqrt{\frac{\lambda F(\omega)(v)}{1+\lambda F(\omega)(v)}}\sum_{u\in N(v)}\sqrt{\frac{\lambda\omega(u)}{1+\lambda\omega(u)}} \leq \sqrt{\frac{\lambda d f(\bar{\omega})}{1+\lambda f(\bar{\omega})}\cdot\frac{\lambda d\bar{\omega}}{1+\lambda\bar{\omega}}},$$

where $d = |N(v)|$ is the degree of vertex v in G and $f(\bar{\omega}) = (1+\lambda\bar{\omega})^{-d}$ is the symmetric version of the recursion (3).

Define $\alpha_{\lambda,d}(x) = \sqrt{\frac{\lambda d f(x)}{1+\lambda f(x)}\cdot\frac{\lambda dx}{1+\lambda x}}$, where as before $f(x) = (1+\lambda x)^{-d}$. The above convexity argument shows that

$$\sqrt{\frac{\lambda F(\omega)(v)}{1+\lambda F(\omega)(v)}}\sum_{u\in N(v)}\sqrt{\frac{\lambda\omega(u)}{1+\lambda\omega(u)}} \leq \alpha_{\lambda,d}(x) \text{ for some } x \in [0,1]. \tag{91}$$

Fixing any λ and d, the critical point of $\alpha_{\lambda,d}(x)$ is achieved at the unique positive $x(\lambda, d)$ satisfying

$$\lambda d x(\lambda,d) = 1 + \lambda f(x(\lambda,d)). \tag{92}$$

It is also easy to verify by checking the derivative $\frac{\mathrm{d}\alpha_{\lambda,d}(x)}{\mathrm{d}x}$ that the maximum of $\alpha_{\lambda,d}(x)$ is achieved at this critical point $x(\lambda,d)$.

Recall that $\hat{x}(\lambda,d)$ is the fixed point satisfying $\hat{x}(\lambda,d) = f(\hat{x}(\lambda,d)) = (1+\lambda\hat{x}(\lambda,d))^{-d}$, and $\alpha(\lambda,d) = \sqrt{\frac{\lambda d\hat{x}(\lambda,d)}{1+\lambda\hat{x}(\lambda,d)}}$. Under the assumption that $\alpha(\lambda,d) \le 1$, we must have $\hat{x}(\lambda,d) \le x(\lambda,d)$. If otherwise $\hat{x}(\lambda,d) > x(\lambda,d)$, then we would have $\lambda d\hat{x}(\lambda,d) > \lambda dx(\lambda,d) = 1+\lambda f(x(\lambda,d)) > 1+\lambda f(\hat{x}(\lambda,d)) = 1+\lambda\hat{x}(\lambda,d)$, contradicting that $\frac{\lambda d\hat{x}(\lambda,d)}{1+\lambda\hat{x}(\lambda,d)} = \alpha(\lambda,d)^2 \le 1$. Therefore, for any $x \in [0,1]$, it holds that

$$\begin{aligned}
\alpha_{\lambda,d}(x) &\le \alpha(d, x(\lambda,d)) \\
&= \sqrt{\frac{\lambda df(x(\lambda,d))}{1+\lambda f(x(\lambda,d))} \cdot \frac{\lambda dx(\lambda,d)}{1+\lambda x(\lambda,d)}} \\
&= \sqrt{\frac{\lambda df(x(\lambda,d))}{1+\lambda x(\lambda,d)}} && \text{(due to (92))} \\
&\le \sqrt{\frac{\lambda df(\hat{x}(\lambda,d))}{1+\lambda\hat{x}(\lambda,d)}} && (\hat{x}(\lambda,d) \le x(\lambda,d)) \\
&= \sqrt{\frac{\lambda d\hat{x}(\lambda,d)}{1+\lambda\hat{x}(\lambda,d)}} \\
&= \alpha(\lambda,d).
\end{aligned}$$

Finally, it is easy to observe that $\alpha(\lambda,d)$ is increasing in d since $\alpha(\lambda,d)$ is increasing in $\hat{x}(\lambda,d)$ and $\hat{x}(\lambda,d)$ is increasing in d. Therefore, $\alpha(\lambda,d) \le \alpha(\lambda,\Delta)$ because $d = |N(v)| \le \Delta$. Combined this with (91), the proposition is proved. ∎

8. Loopy BP: Proof of Theorem 3. Consider the version of loopy BP defined with the following sequence of messages: For all $t \ge 1$, for $v \in V$,

$$\tilde{R}_v^t = \lambda \prod_{w \in N(v)} \frac{1}{1+\tilde{R}_w^{t-1}}. \tag{93}$$

The system of equations specified by (93) is equivalent to the one in (3) in the following sense: Given any set of initial messages $(\tilde{R}_v^0)_{v\in V} \in \mathbb{R}_{\ge 0}$, it holds that $\tilde{R}_v^t = \lambda F^t(\bar{\omega})(v)$ for appropriate $\bar{\omega}$ which depends on the initial messages, i.e., $(\tilde{R}_v^0)_{v\in V}$. F^t is the tth iteration of the function F.

Of interest is the quantity $q^t(v)$, $v \in V$, defined as follows:

$$\tilde{q}^t(v) = \frac{\tilde{R}_v^t}{1+\tilde{R}_v^t}.$$

From Lemma 4, there exists $\tilde{q}^* \in [0,1]^V$ such that $\tilde{q}^t$ converges to $\tilde{q}^*$ as $t \to \infty$, in the sense that $\tilde{q}^t/\tilde{q}^* \to 1$. It is elementary to show that the following holds for any $t > 0$, any $p \in V$, and $v \in N(p)$:

$$\begin{aligned}
\frac{q^t(v,p)}{\mu(v \text{ is occupied} \mid p \text{ is unoccupied})} &= \frac{q^t(v,p)}{q^*(v,p)} \frac{q^*(v,p)}{\tilde{q}^*(v)} \cdot \frac{\tilde{q}^*(v)}{\mu(v \text{ is occupied})} \\
&\quad \cdot \frac{\mu(v \text{ is occupied})}{\mu(v \text{ is occupied} \mid p \text{ is unoccupied})}.
\end{aligned}$$

The theorem follows by showing that each of the four ratios on the r.h.s. are sufficiently close to 1. For the first two ratios we use Theorem 25, and for the third one we use the Lemma 26.

THEOREM 25. *For all $\delta, \epsilon > 0$, there exists $\Delta_0 = \Delta_0(\delta, \epsilon)$ and $C = C(\delta, \epsilon)$, such that for all $\Delta \geq \Delta_0$, all $\lambda < (1-\delta)\lambda_c(\Delta)$, all graphs G of maximum degree Δ and girth ≥ 6, and all $\epsilon > 0$ the following holds:*

There exists $q^* \in [0,1]^E$ such that for $t \geq C$, for all $p \in V$, $v \in N(p)$ we have that

$$\left|\frac{q^t(v,p)}{q^*(v,p)} - 1\right| \leq \epsilon \quad \textit{and} \quad \left|\frac{q^*(v,p)}{\tilde{q}^*(v)} - 1\right| \leq \epsilon, \tag{94}$$

where $q^t(v,p)$ is defined in (1).

Proof. Note that by denoting $\omega^t(v,p) = \frac{R^t_{p\to v}}{\lambda}$, we have

$$\omega^{t+1}(v,p) = H(\omega^t)(v,p),$$

where H is as defined in (8). Then the convergence of $q^t(v,p) = \frac{R^t_{p\to v}}{1+R^t_{p\to v}}$ to a unique fixed point q^* follows from Corollary 6. More precisely, there is $\Delta_0 = \Delta_0(\delta)$ and $C = C(\epsilon_0, \delta)$ such that for all $\Delta > \Delta_0$ all $\lambda < (1-\delta)\lambda_c(T_\Delta)$ and all $t > C$,

$$\left|\omega^t(v,p) - \omega^*(v,p)\right| \leq \epsilon_0.$$

Note that for all $t > 1$, we have $\omega^t(v,p), \omega^*(v,p) \in [(1+\lambda)^{-\Delta}, 1]$, where $(1+\lambda)^{-\Delta} > 1/36$ for $\lambda < \lambda_c(T_\Delta)$ for all sufficiently large Δ. Then

$$\left|\frac{q^t(v,p)}{q^*(v,p)} - 1\right| = \left|\frac{\omega^t(v,p)}{\omega^*(v,p)} \cdot \frac{1+\omega^*(v,p)}{1+\omega^t(v,p)} - 1\right| = \frac{|\omega^t(v,p) - \omega^*(v,p)|}{\omega^*(v,p)(1+\omega^t(v,p))} \leq 36\epsilon_0.$$

By choosing $\epsilon_0 = \frac{\epsilon}{36}$, we have $|\frac{q^t(v,p)}{q^*(v,p)} - 1| \leq \epsilon$.

We then show that there is a $\Delta_0 = O(\frac{1}{\delta\epsilon})$ such that for all $\Delta > \Delta_0$ and all $\lambda < (1-\delta)\lambda_c(T_\Delta)$, the fixed points of the two BPs have $|\frac{q^*(v,p)}{\tilde{q}^*(v)} - 1| \leq \epsilon$.

Let $\omega^t(v,p) = \frac{q^t(v,p)}{\lambda(1-q^t(v,p))}$ and $\tilde{\omega}^t(v) = \frac{\tilde{q}^t(v)}{\lambda(1-\tilde{q}^t(v))}$. It follows that

$$\omega^{t+1}(v,p) = \prod_{u\in N(v)\setminus\{p\}} \frac{1}{1+\lambda\omega^t(u,v)} = (1+\lambda\omega^t(p,v)) \prod_{u\in N(v)} \frac{1}{1+\lambda\omega^t(u,v)},$$

$$\tilde{\omega}^{t+1}(v) = \prod_{u\in N(v)} \frac{1}{1+\lambda\tilde{\omega}^t(u)}.$$

We also define

$$\omega^{t+1}(v) = \prod_{u\in N(v)} \frac{1}{1+\lambda\omega^t(u,v)},$$

and therefore $\omega^{t+1}(v,p) = (1+\lambda\omega^t(p,v))\omega^{t+1}(v)$. Note that $\omega^t(p,v) \in (0,1]$, and thus $|\omega^{t+1}(v,p) - \omega^{t+1}(v)| \leq \lambda$. Also recall that $\lambda < \lambda_c(T_\Delta) \leq 3/(\Delta-2)$ for all sufficiently large Δ, and therefore

$$\left|\omega^{t+1}(v,p) - \omega^{t+1}(v)\right| \leq 3/(\Delta-2).$$

Let $\Psi(\cdot)$ be as defined in (5). Note for $t > 1$ both $\omega^{t+1}(v,p)$ and $\omega^{t+1}(v)$ are from the range $[(1+\lambda)^{-\Delta}, 1]$. By (6), for $\lambda < \lambda_c(T_\Delta)$ for all sufficiently large Δ, and we have

$$\left|\Psi\left(\omega^{t+1}(v,p)\right) - \Psi\left(\omega^{t+1}(v)\right)\right| \le 9/(\Delta-2). \tag{95}$$

We assume that $|\Psi(\omega^t(v,p)) - \Psi(\tilde\omega^t(v))| \le \epsilon_0$ for all $(v,p) \in E$. Then due to (11),

$$\left|\Psi\left(\omega^{t+1}(v)\right) - \Psi\left(\tilde\omega^{t+1}(v)\right)\right| \le (1-\delta/6) \cdot \max_{u\in N(v)} \left|\Psi\left(\omega^t(u,v)\right) - \Psi\left(\tilde\omega^t(u)\right)\right| \le (1-\delta/6)\epsilon_0.$$

Combined with (95), by triangle inequality, we have

$$\left|\Psi\left(\omega^{t+1}(v,p)\right) - \Psi\left(\tilde\omega^{t+1}(v)\right)\right| \le (1-\delta/6)\epsilon_0 + 9/(\Delta-2),$$

which is at most ϵ_0 as long as $\Delta \ge \Delta_0 \ge \frac{54}{\delta\epsilon_0} + 2$. It means that if $|\Psi(\omega^t(v)) - \Psi(\omega^t(v,p))| \le \epsilon_0 \le \frac{54}{\delta(\Delta_0-2)}$, then $|\Psi(\omega^{t+1}(v,p)) - \Psi(\tilde\omega^{t+1}(v))| \le \frac{54}{\delta(\Delta_0-2)}$. Knowing the convergences of $\omega^t(v,p)$ to $\omega^*(v,p)$ and $\tilde\omega^t(v)$ to $\omega^*(v)$ as $t\to\infty$, this gives us that

$$\left|\Psi\left(\omega^*(v,p)\right) - \Psi\left(\tilde\omega^*(v)\right)\right| \le \frac{54}{\delta(\Delta_0-2)}.$$

By (6), it implies $|\omega^*(v,p) - \tilde\omega^*(v)| \le \frac{162}{\delta(\Delta_0-2)}$. Again since $\omega^*(v,p), \tilde\omega^*(v) \in [1/36, 1]$ when $\lambda < \lambda_c(T_\Delta)$ for sufficiently large Δ. It holds that

$$\left|\frac{q^*(v,p)}{\tilde q^*(v)} - 1\right| = \left|\frac{\omega^*(v,p)}{\tilde\omega^*(v)} \cdot \frac{1+\lambda\tilde\omega^*(v)}{1+\lambda\omega^*(v,p)} - 1\right| \le \frac{6000}{\delta(\Delta_0-2)}.$$

By choosing a suitable $\Delta_0 = O(\frac{1}{\delta\epsilon})$, we can make this error bounded by ϵ. ∎

LEMMA 26. *For all $\delta, \epsilon > 0$, there exists $\Delta_0 = \Delta_0(\delta,\epsilon)$ and $C = C(\delta,\epsilon)$, such that for all $\Delta \ge \Delta_0$, all $\lambda < (1-\delta)\lambda_c(\Delta)$, all graphs G of maximum degree Δ and girth ≥ 6, the following holds: Let $\mu(\cdot)$ be the Gibbs distribution, and for all $v \in V$ we have*

$$\left|\frac{\tilde q^*(v)}{\mu(v \textit{ is occupied})} - 1\right| \le \epsilon.$$

Proof. It holds that

$$\left|\frac{\tilde q^*(v)}{\mu(v \text{ occupied})} - 1\right| = \left|\frac{q^*(v)}{\frac{\lambda}{1+\lambda}\mathbb{E}\left[\mathbf{R}(X,v)\right]} \frac{\frac{\lambda}{1+\lambda}\mathbb{E}\left[\mathbf{R}(X,v)\right]}{\mu(v \text{ occupied})} - 1\right|, \tag{96}$$

where the expectation in the nominator is w.r.t. the random variable X which is distributed as in μ. For showing the lemma we need to bound appropriately the two ratios on the r.h.s. of (96). For this we use the following two results. The first one is that

$$\left|\frac{\frac{\lambda}{1+\lambda}\mathbb{E}\left[\mathbf{R}(X,v)\right]}{\mu(v \text{ is occupied})} - 1\right| \le 200e^e\lambda. \tag{97}$$

Before proving (97), let us show how it implies the lemma, together with Lemma 13. For any independent set σ and any v, it holds that $e^{-e} \le \omega^*(v), \mathbf{R}(\sigma,v) \le 1$. Then, Lemma 13 implies that

$$\left|\frac{\omega^*(v)}{\mathbb{E}\left[\mathbf{R}(X,v)\right]} - 1\right| \le \epsilon/20. \tag{98}$$

Noting that by definition it holds that $\tilde{q}^*(v) = \frac{\lambda\omega^*}{1+\lambda\omega^*}$, and we have that

$$\left|\frac{\tilde{q}^*(v)}{\frac{\lambda}{1+\lambda}\mathbb{E}\left[\mathbf{R}(X,v)\right]} - 1\right| = \left|\frac{1+\lambda}{1+\lambda\omega^*(v)}\frac{\omega^*(v)}{\mathbb{E}\left[\mathbf{R}(X,v)\right]} - 1\right|$$

$$\leq \frac{10\lambda}{(1+\lambda\omega^*(v))}\frac{\omega^*(v)}{\mathbb{E}\left[\mathbf{R}(X,v)\right]} + \left|\frac{\omega^*(v)}{\mathbb{E}\left[\mathbf{R}(X,v)\right]} - 1\right| \leq \epsilon/15. \tag{99}$$

In the last inequality we use (98), the fact that $\lambda < 2e/\Delta$ and Δ is sufficiently large. The lemma follows by plugging (97) and (99) into (96). We proceed by showing (97). It holds that

$$\mu(v \text{ is occupied}) = \frac{\lambda}{1+\lambda}\mu(v \text{ is unblocked}) \tag{100}$$

We are going to express $\mu(v \text{ is unblocked})$ it terms of the quantity $\mathbf{R}(\cdot,\cdot)$. For X distributed as in μ it is elementary to verify that

$$\mathbb{E}\left[\mathbf{R}(X,v) \mid v \text{ is unoccupied}\right] = \mu(v \text{ is unblocked} | v \text{ is unoccupied}). \tag{101}$$

Furthermore, it holds that

$$\begin{aligned}
&\mathbb{E}\left[\mathbf{R}(X,v)\right] \\
&\quad = \mu(v \text{ occupied}) \cdot \mathbb{E}\left[\mathbf{R}(X,v) \mid v \text{ occupied}\right] + \mu(v \text{ unoccupied}) \\
&\qquad \cdot \mathbb{E}\left[\mathbf{R}(X,v) \mid w \text{ unoccupied}\right] \\
&\quad \leq \mu(v \text{ occupied}) + \mathbb{E}\left[\mathbf{R}(X,v) \mid v \text{ unoccupied}\right] && (\text{since } 0 < R(X,v) \leq 1) \\
&\quad \leq 2\lambda + \mathbb{E}\left[\mathbf{R}(X,v) \mid v \text{ unoccupied}\right] && (\text{since } \mu(v \text{ occupied}) \leq 2\lambda).
\end{aligned}$$

Since $e^{-e} \leq \mathbf{R}(X,v) \leq 1$, the inequality above yields

$$\mathbb{E}\left[\mathbf{R}(X,v) \mid v \text{ unoccupied}\right] \geq (1 - 2e^e\lambda)\,\mathbb{E}\left[\mathbf{R}(X,v)\right].$$

Also, using the fact that $\mathbf{R}(X,v) > 0$, we get

$$\mathbb{E}\left[\mathbf{R}(X,v) \mid v \text{ unoccupied}\right] \leq \frac{\mathbb{E}\left[\mathbf{R}(X,v)\right]}{\mu(v \text{ is unoccupied})} \leq (1+5\lambda)\mathbb{E}\left[\mathbf{R}(X,v)\right].$$

In the last inequality we use the fact that $\mu(w \text{ is occupied}) \leq 2\lambda$. From the above two inequalities we get that

$$\left|\mathbb{E}\left[\mathbf{R}(X,w) \mid w \text{ unoccupied}\right] - \mathbb{E}\left[\mathbf{R}(X,w)\right]\right| \leq 10e^e\lambda. \tag{102}$$

In a very similar manner as above, we also get that

$$\left|\mu(v \text{ is unblocked} | v \text{ is unoccupied}) - \mu(v \text{ is unblocked})\right| \leq 10e^e\lambda. \tag{103}$$

Combining (101), (102), (103), (100) and using the fact that $e^{-e} \leq \mu(v \text{ is unblocked})$, $\mathbb{E}\left[\mathbf{R}(X,w)\right]$ we get the following:

$$\mu(v \text{ is occupied}) = \frac{\lambda}{1+\lambda}\mathbb{E}\left[\mathbf{R}(X,w)\right](1 + 50e^e\lambda). \tag{104}$$

Then (97) follows from (104). ∎

The theorem follows by showing that

$$\left|\frac{\mu(v \text{ is occupied})}{\mu(v \text{ is occupied} \mid p \text{ is unoccupied})} - 1\right| \leq 10/\Delta.$$

From Bayes' rule we get that $\mu(v \text{ is occupied} \mid p \text{ is unoccupied}) = \frac{\mu(v \text{ is occupied})}{\mu(p \text{ is unoccupied})}$. Using this observation we get that

$$\left|\frac{\mu(v \text{ is occupied})}{\mu(v \text{ is occupied} \mid p \text{ is unoccupied})} - 1\right| = |\mu(p \text{ is unoccupied}) - 1| \leq 10/\Delta.$$

In the last inequality we use the fact that $0 \leq \mu(p \text{ is occupied}) \leq \lambda$.

9. Basic properties of Glauber dynamics.

9.1. Continuous versus discrete time chains. For many of our results we have a simpler proof when instead of a discrete time Markov chain we consider a continuous time version of the chain. That is, consider the Glauber dynamics where the spin of each vertex is updated according to an independent Poisson clock with rate $1/n$.

We use the following observation [24, Corollary 5.9] as a generic tool to argue that typical properties of continuous time chains are typical properties of the discrete time chains too.

Observation 27. Let (X_t) by any discrete time Markov chain on state space Ω, and let (Y_t) be the corresponding continuous-time chain. Then for any property $P \subset \Omega$ and positive integer t, we have that

$$\mathbf{Pr}\left[X_t \notin P\right] \leq e\sqrt{t}\mathbf{Pr}\left[Y_t \notin P\right].$$

Observation 27 would suffice for our purposes when $\Delta = \Omega(\log n)$, but not for Glauber dynamics on graphs of, e.g., constant degree. For the latter case, instead of focusing on specific times t in discrete time, our goal will be to show how events which are rare at a single instant in continuous time must also be rare over a time interval of length $O(n)$ in discrete time, without taking a union bound over all the times in the time interval.

Let the set Ω contain all the independent sets of G. We say that a function $f : \Omega \to \mathbb{R}$ has "total influence" J if for every independent set $X \in \Omega$ we have

$$\mathbb{E}\left[|f(X') - f(X)|\right] \leq J/n,$$

where X' is the result of one Glauber dynamics update, starting from X.

The next result [10, Lemma 13] shows that, for functions f which have Lipschitz constant $O(1/\Delta)$ and total influence $J = O(1)$, in order to prove high-probability bounds for the discrete-time chain that apply for all times in an interval of length $O(n)$, it suffices to be able to prove a similar bound at a single instant in continuous time.

LEMMA 28 (Hayes [10]). *Suppose $f : \Omega \to \mathbb{R}$ is a function of independent sets of G and f has Lipschitz constant $\alpha < O(1/\Delta)$ and total influence $J = O(1)$. Let $X_0 = Y_0$ be given and let $(X_t)_{t\geq 0}$ be continuous-time single site dynamics on the hard-core model of G and let $(Y_i)_{i=0,1,2...}$ be the corresponding discrete-time dynamics.*

Suppose that t_0 is a positive integer and S is a measurable set of real numbers, such that for all $t \geq t_0$, $\mathbf{Pr}\left[f(X_t) \in S\right] \geq 1 - \exp(-\Omega(\Delta))$. Then, for all $\epsilon \in \Omega(1)$

and all integers $t_1 \geq t_0$, where $t_1 - t_0 = O(n)$, we have that

$$\mathbf{Pr}\left[(\forall i \in \{t_0, t_0+1, \ldots, t_1\})\ f(Y_i) \in S \pm \epsilon\right] \geq 1 - \exp\left(-\Omega(\Delta)\right),$$

where the hidden constant in Ω notation depends only on the hidden constant in the assumption.

9.2. G versus G^* and comparison. Consider G with girth ≥ 7. For such a graph and some vertex w in G, the radius 3 ball around w is a tree. We let G^*_w be the graph that is derived from G by orienting toward w every edge that is within distance 2 from w. (An edge $\{w_1, w_2\} \in E$ is at distance ℓ from w if the minimum distance between w and any of w_1, w_2 is ℓ.) For a vertex $x \in G^*_w$, we let $N^*(x) \subseteq N(x)$ contain every z in the neighborhood of x such that either the edge between x, z is unoriented or the orientation is toward x.

We let the Glauber dynamics (X^*_t) on the hard-core model with underlying graph G^*_w and fugacity λ be a Markov chain whose transition $X_t \to X_{t+1}$ is defined by the following:

1. Choose u uniformly at random from V.
2. If $N^*(u) \cap X^*_t = \emptyset$, then let

$$X^*_{t+1} = \begin{cases} X^*_t \cup \{u\} & \text{with probability } \lambda/(1+\lambda), \\ X^*_t \setminus \{u\} & \text{with probability } 1/(1+\lambda). \end{cases}$$

3. If $N^*(w) \cap X_t \neq \emptyset$, then let $X^*_{t+1} = X^*_t$.

The state space of (X^*_t) that is implied by the above is a superset of the independent sets of G, since there are pairs of vertices which are adjacent in G while they can both be occupied in X^*_t.

The motivation for using G^*_w and (X^*_t) is better illustrated by considering Lemma 12. In Lemma 12 we establish a recursive relation for $\mathbf{R}()$ for G of girth ≥ 6, in the setting of the Gibbs distribution. An important ingredient in the proof there is that for every vertex x conditioned on the configuration at x and the vertices at distance ≥ 3 from x, the children of x are mutually independent of each other under the Gibbs distribution.

For establishing the uniformity property for Glauber dynamics we need to establish a similar "conditional independence" relation but in the dynamic setting of Markov chains. To obtain this, we will need that G has girth at least 7. Clearly, the conditional independence of Gibbs distribution no longer holds for the Glauber dynamics. To this end we employ the following: Instead of considering G and the standard Glauber dynamics (X_t), we consider G^*_w and the corresponding dynamics (X^*_t).

Using G^*_w and (X^*_t) we get (in the dynamics setting) an effect which is similar to the conditional independence. During the evolution of (X^*_t) the neighbors of w can only exchange information through paths of G^*_w which travel outside the ball of radius 3 around w, i.e., $B_3(w)$. This holds due to the girth assumption for G^*_w and the definition of (X^*_t). In turn this implies that conditional on the configuration of X^*_t outside $B_3(w)$, the (grand)children of w are mutually independent.

The above trick allows us to get a recursive relation for $R(X^*_t, w)$ similar to that for the Gibbs distribution. To argue that a somehow similar relation holds for the standard dynamics (X_t), we use the following result which states that if (X^*_t) and (X_t) start from the same configuration, then after $O(n)$ the number of disagreements between the two chains is not too large.

LEMMA 29. *For $\gamma > 0$, $C_1 > 0$, there exists Δ_0, $C_2 > 0$ such that the following is true: For $w \in V$ consider G^*_w of maximum degree $\Delta > \Delta_0$ and girth at least 7. Also, let (X_t) and (X^*_t) be the continuous time Glauber dynamics on the hard-core model with fugacity $\lambda < (1-\delta)\lambda_c(\Delta)$, underlying graphs G and G^*_u, respectively.*

Assume that (X^*_t) and (X_t) are maximally coupled. Then, if $X_0 = X^*_0$ for X_0 which $(400, R)$-nice at w for radius $R \le \Delta^{9/10}$, we have that

$$\Pr[\forall s \le C_1 n,\ \forall u \in V\ |(X_s \oplus X^*_s) \cap B_2(u)| \le \gamma\Delta] \ge 1 - \exp\left(-\Delta/C_2\right).$$

Before proving Lemma 29 we need to introduce certain notions.

Let us call Z a "generalized Poisson random variable with jumps α and instantaneous rate $r(t)$" if Z is the result of a continuous-time adapted process, which begins at 0 and in each subsequent infinitesimal time interval, samples an increment ∂Z from some distribution over $[0, \alpha]$, having mean $\le r(t)dt$. Z, the sum of the increments over all times $0 < t < 1$, is a random variable, as is the maximum observed rate, $r^* = \max_{t \in [0,1]} r(t)$.

Remark 30. In the special case where $\alpha \ge 1$ and the distribution is supported in $\{0, 1\}$ with constant rate $\mu \cdot dt$, Z is a Poisson random variable with mean μ.

We are going to use the following result [10, Lemma 12].

LEMMA 31 (Hayes). *Suppose Z is a generalized Poisson random variable with maximum jumps α and maximum observed rate r^*. Then, for every $\mu > 0$, $C > 1$ it holds that*

$$\mathbf{Pr}\left[Z \ge C\mu \text{ and } r^* \le \mu\right] \le \exp\left[-\frac{\mu}{\alpha}(C\ln(C) - C + 1)\right] < \left(\frac{e}{C}\right)^{\mu C \alpha}.$$

Proof of Lemma 29. In this proof assume that γC_1 is sufficiently small constant. Also, let $D = \cup_{t \le C_1 n}(X_t \oplus X^*_t)$, i.e., D denotes the set of all vertices which are disagreeing at least once during the time interval from 0 to $C_1 n$. Given some vertex $u \in V$ let $A_u = \cup_{t \le C_1 n} X_t \cap N(u)$ and $A^*_u = \cap_{t \le C_1 n} X^*_t \cap N(u)$. That is, A_u contains every $z \in N(u)$ for which there exists at least one $s < C_1 n$ such that z is occupied in X_s, and similarly for A^*_u. Finally, let the integer $r = \lfloor \gamma^5 \frac{\Delta}{\log \Delta} \rfloor$.

Let $\mathcal{A}$ denote the event that $\exists s \le C_1 n,\ \exists u \in V\ |(X_s \oplus X^*_s) \cap S_2(u)| \ge \gamma\Delta/2$. Consider the events $\mathcal{B}_1, \mathcal{B}_2$, $\mathcal{B}_3$, $\mathcal{B}_4$, and $\mathcal{B}_5$ to be defined as follows: $\mathcal{B}_1$ denotes the event that $D \nsubseteq B_r(w)$. $\mathcal{B}_2$ denotes the event that $|D| \ge \gamma^3 \Delta^2$. $\mathcal{B}_3$ denotes the event that the total number of disagreements that appear in $N(u)$, for every $u \in V$, is at most $\gamma^3\Delta$. Finally, $\mathcal{B}_4$ denotes the event that there exists $u \in B_{100}(w)$ such that either $|A(u)| \ge \gamma^3\Delta$ or $|A^*(u)| \ge \gamma^3\Delta$.

Then, the lemma follows by noting the following:

$$\mathbf{Pr}\left[\exists s \le C_1 n,\ \exists u \in V\ |(X_s \oplus X^*_s) \cap B_2(u)| \ge \gamma\Delta\right] \le \mathbf{Pr}\left[\mathcal{A}\right] + \mathbf{Pr}\left[\mathcal{B}_3\right]. \tag{105}$$

The lemma follows by bounding appropriately the probability terms on the r.h.s. of (105).

First consider $\mathbf{Pr}[\mathcal{A}]$. Let $\mathcal{B} = \mathcal{B}_1 \cup \mathcal{B}_2 \cup \mathcal{B}_3 \cup \mathcal{B}_4$. We bound $\mathbf{Pr}[\mathcal{A}]$ by using $\mathcal{B}$ as follows:

$$\begin{aligned}
\mathbf{Pr}\left[\mathcal{A}\right] &= \mathbf{Pr}\left[\mathcal{B}, \mathcal{A}\right] + \mathbf{Pr}\left[\bar{\mathcal{B}}, \mathcal{A}\right] \\
&\le \mathbf{Pr}\left[\mathcal{B}\right] + \mathbf{Pr}\left[\bar{\mathcal{B}}, \mathcal{A}\right] \\
&\le \sum_{i=1}^{4} \mathbf{Pr}\left[\mathcal{B}_i\right] + \mathbf{Pr}\left[\bar{\mathcal{B}}, \mathcal{A}\right], \qquad (106)
\end{aligned}$$

where the last inequality follows by applying a simple union bound.

Consider some vertex $u \in V$ and let Z be the total number of disagreements that ever occur in $S_2(u)$ up ot the first time that either $\mathcal{B}$ occurs or up to time $C_1 n$, whichever happens first. If $u \notin B_r(w)$, then Z is always zero since we stop the clock when $D \not\subseteq B_{r-1}(w)$. So our focus is on the case where $u \in B_{r-1}(w)$. For such u the random variable Z follows a generalized Poisson distribution, with jumps of size 1 and maximum observed rate at most $30\gamma^3 \Delta dt/n$, over at most $C_1 n$ time units. To see this consider the following.

Given that $\mathcal{B}$ does not occur, disagreements in $S_2(u)$ may be caused due to the following categories of disagreeing edges. Each disagreement in $N(u)$ has at most $\Delta - 1$ disagreeing edges in $S_2(u)$. Since the number of disagreements that appear in $N(u)$ during the time period up to $C_1 n$ is at most $\gamma^3 \Delta$, there are at most $\gamma^3 \Delta^2$ disagreeing edges incident to $S_2(u)$. On the whole there are at most $\gamma^3 \Delta^2$ disagreements from vertices different than those in $N(u)$. Each one of them has at most one neighbor in $S_2(u)$, since the girth is at least 7. That is, there are additional $\gamma^3 \Delta^2$ many disagreeing edges. Finally, disagreements on $S_2(u)$ may be caused by edges which belong to $G \oplus G^*_w$. There are at most Δ^3 many such edges. Each one of these edges generates disagreements only on the vertex on its tail. Since the out-degree in G^*_w is at most 1, there are Δ^2 disagreeing edges from $G \oplus G^*_w$ which are incident to $S_2(v)$. Additionally, each one of these edges should point to an occupied vertex so as to be disagreeing. Since $\mathcal{B}_4$ does not occur, there at most $2\gamma^3 \Delta^2$ edges in $G \oplus G^*_w$ which point to an occupied vertex and have the tail in S_2.

From the above observations, we have that there are at most $10\gamma^3 \Delta^2$ disagreeing edges incident to S_2. For the new disagreement to occur in S_2 due to a given such edge, a specific vertex must chosen and should become occupied, which occurs with rate at most $e \cdot dt/(n\Delta)$.

Using Lemma 31, applied with $\mu = 30 C_1 \gamma^3 \Delta$, $\alpha = 1$, and $C = \gamma\Delta/\mu$, we have that

$$\mathbf{Pr}\left[Z \geq \gamma\Delta\right] \leq \left(30 e \gamma^2 C_1\right)^{\gamma\Delta}.$$

Taking a union bound over the, at most, Δ^r vertices in $B_r(v)$, we get that

$$\mathbf{Pr}\left[\bar{\mathcal{B}}, \mathcal{A}\right] \leq \Delta^r \left(30 e \gamma^2 C_1\right)^{\gamma\Delta} = \exp\left(-\Delta/C_3\right), \tag{107}$$

where $C_3 = C_3(\gamma) > 0$ is a sufficiently large number. In the last derivation we used the fact that $r \leq \frac{\gamma^5 \Delta}{\log \Delta}$.

We proceed by bounding the probability of the events $\mathcal{B}_1$, $\mathcal{B}_2$, $\mathcal{B}_3$, and $\mathcal{B}_4$. The approach is very similar to the proof of [10, Theorem 27]. We repeat it for the sake of completeness.

Recall that $\mathcal{B}_1$ denotes the event that $D \not\subseteq B_r(w)$. The bound for $\mathbf{Pr}[\mathcal{B}_1]$ uses standard arguments of disagreement percolation. First we observe that every disagreement outside $B_r(w)$ must arise via some path of disagreement which starts within $B_2(w)$. That is, we need at least one path of disagreement of length $r-4$. We fix a particular path of length $r-4$ with $B_r(w)$. Let us call it $\mathcal{P}$. We are going to bound the probability that disagreements percolate along $\mathcal{P}$ within $C_1 n$ time units. Let us call this probability ρ.

The number of steps along this path that a disagreement actually percolates is a generalized Poisson random variable with jumps 1 and maximum overall rate at most $C_1 e/\Delta$. This follows by noting that the maximum instantaneous rate is at most $e \cdot dt/(n\Delta)$ integrated over $C_1 n$ time units. We use Lemma 31 to bound the probability for the disagreement to percolate along $\mathcal{P}$, i.e., ρ. Setting $\mu = eC_1/\Delta$, $\alpha = 1$, and

$C = (r-4)/\mu$ in Lemma 31 yields the following bound for ρ:

$$\rho \leq \left(\frac{e^2 C_1}{\Delta(r-4)}\right)^{r-4}.$$

The above bound holds for any path of length $r-4$ in $B_r(w)$. Taking a union bound over the at most Δ^3 starting point in $B_2(w)$ and the at most Δ^{r-4} paths of length $r-4$ from a given starting point we get that

$$\mathbf{Pr}\left[\mathcal{B}_1\right] \leq \Delta^3 \left(\frac{e^2 C_1}{r-4}\right)^{r-4} \leq \exp\left(-\Delta/C_4\right), \tag{108}$$

where $C_4 = C_4(\gamma) > 0$ is a sufficiently large number.

Recall that $\mathcal{B}_2$ denotes the event that $|D| \geq \gamma^3\Delta^2$. For $\Pr[\mathcal{B}_2]$ we consider the waiting time τ_i for the ith disagreement, counting from when the $(i-1)$st disagreement is formed. The event $\mathcal{B}_2$ is equivalent to $\sum_{i=1}^{(\gamma^3\Delta^2)} \tau_i \leq C_1 n$.

Each new disagreement can be attributed to either an edge joining it to an existing disagreement or to one of the edges in $G \oplus G_w^*$. It follows easily that the total number of such edges is at most $|G \oplus G_w^*| + |(i-1)\Delta| = \Delta^3 + (i-1)\Delta$. Furthermore, for the new disagreement to occur due to a given such edge, a specific vertex must chosen, which occurs with rate at most $e \cdot dt/(n\Delta)$.

The above observations suggest that the waiting time τ_i is stochastically dominated by an exponential distribution with mean $n/[e(\Delta^2+i-1)]$, even conditioning on an arbitrary previous histry $\tau_1, \tau_2, \ldots, \tau_{i-1}$. Therefore, $\sum_i \tau_i$ is stochastically dominated by the sum of independent exponential distributions with mean $n/[e(\Delta^2+i-1)]$.

Applying [10, Corollary 26] to $\tau_1 + \cdots + \tau_{(\gamma^3\Delta^2)}$ with

$$\mu = \sum_{i=1}^{(\gamma^3\Delta^2)} \frac{n}{e(\Delta^2+i-1)} \geq \int_0^{(\gamma^3\Delta^2)} \frac{n}{e(\Delta^2+x)} dx = \frac{n}{e}\log\left(1+\gamma^3\right)$$

and

$$V = \sum_{i=1}^{(\gamma^3\Delta^2)} \frac{n^2}{e^2(\Delta^2+i-1)^2} \leq \int_0^{\infty} \frac{n^2}{e^2(\Delta^2+x-1)^2} dx = \frac{n^2}{e^2(\Delta^2-1)},$$

all the above yield

$$\Pr[\mathcal{B}_2] \leq \exp\left(-(\mu - C_1 n)^2/(2V)\right) \leq \exp\left(-\Delta^2/C_5\right), \tag{109}$$

where $C_5 = C_5(\gamma) > 0$ is sufficiently large number.

Let Y be the total number of disagreements that ever occur in $N(u)$ up to the first time that either $D \not\subseteq B_{r-1}(w)$ or $|D| > \gamma^3\Delta^2$ occur or time $C_1 n$, whichever happens first. The variable Y follows a generalized Poisson distribution with jumps of size 1. It is direct to check that the maximum observed rate is at most $(\gamma^3\Delta^2+2\Delta)e\cdot dt/(\Delta n) \leq 10\gamma^3\Delta dt/n$, integrated over at most $C_1 n$ time units. This is because the clock stops when $|D| \geq \gamma^3\Delta^2$ and since G has girth at least 7 it is only vertex u that is adjacent to more than one element of $N(u)$. Hence there are at most $\gamma^3\Delta^2 + \Delta - 1$ edges joining joining a disagreement with some vertex in $N(u)$ before the clock stops. Furthermore, disagreements on $N(u)$ may also be caused by incident edges which belong to $G \oplus G_w^*$.

Each vertex in $v \in N(u)$ is incident to at most one edge which belongs to $G \oplus G_w^*$ and could cause disagreement in v. That is, $N(u)$ has at most at most Δ such edges.

Applying Lemma 31, once more, for Y with $\mu = 10C_1\gamma^3\Delta$, $\alpha = 1$, and $C = \gamma^{3/2}\Delta/\mu$ we get that

$$\mathbf{Pr}\left[Y \geq \gamma^2\Delta\right] \leq \left(\frac{10eC_1\gamma^3\Delta}{\gamma^{3/2}\Delta}\right)^{\gamma^{3/2}\Delta} \leq \left(10eC_1\gamma^{3/2}\right)^{\gamma^{3/2}\Delta}.$$

Taking a union bound over the at most Δ^r vertices in $B_r(w)$ gives an upper bound for the probability the event $\mathcal{B}_3$ happens and at the same time neither $\mathcal{B}_1$ nor $\mathcal{B}_2$ occur. That is,

$$\mathbf{Pr}\left[\bar{\mathcal{B}}_1 \text{ and } \bar{\mathcal{B}}_2 \text{ and } \mathcal{B}_3\right] \leq \Delta^r\left(10eC_1\gamma^{3/2}\right)^{\gamma^{3/2}\Delta}. \tag{110}$$

Letting $\mathcal{C} = \mathcal{B}_1 \cup \mathcal{B}_2$, we have that

$$\begin{aligned}
\mathbf{Pr}\left[\mathcal{B}_3\right] &= \mathbf{Pr}\left[\mathcal{C}, \mathcal{B}_3\right] + \mathbf{Pr}\left[\bar{\mathcal{C}}, \mathcal{B}_3\right] \\
&\leq \mathbf{Pr}\left[\mathcal{C}\right] + \mathbf{Pr}\left[\bar{\mathcal{C}}, \mathcal{B}_3\right] \\
&\leq \mathbf{Pr}\left[\mathcal{B}_1\right] + \mathbf{Pr}\left[\mathcal{B}_2\right] + \mathbf{Pr}\left[\bar{\mathcal{B}}_1 \text{ and } \bar{\mathcal{B}}_2 \text{ and } \mathcal{B}_3\right] \\
&\qquad\qquad \text{(by a union bound for } \mathbf{Pr}\left[\mathcal{C}\right]\text{)} \\
&\leq \exp\left(-\Delta/C_6\right), \qquad (111)
\end{aligned}$$

where $C_6 = C_6(\gamma) > 0$. In the last inequality we used (110), (109), and (108).

As far as $\mathbf{Pr}\left[\mathcal{B}_4\right]$ is concerned, first recall that $\mathcal{B}_4$ denotes the event that there exists $z \in B_{100}(w)$ such that either $|A(u)| \geq \gamma^3\Delta$ or $|A^*(u)| \geq \gamma^3\Delta$. Fix some vertex $z \in B_{100}(w)$. Without loss of generality we consider the chain X_t. There are two cases for z. The first one is that z is occupied in X_0. The second one is z is not occupied in X_0. Then the probability that the vertex z is updated and becomes occupied at least once up to time $C_1 n$ is at most $2C_1e/\Delta$, regardless of the rest of the vertices.

Fix some vertex $u \in B_{100}(w)$. Let J_u be the number of vertices $z \in N(u)$ which are *unoccupied* in X_0, but they get into A_u. J_u is dominated by the binomial distribution with parameters Δ and $2C_1e/\Delta$, i.e., $\mathcal{B}(\Delta, 2C_1e/\Delta)$. Using Chernoff's bounds we get that

$$\mathbf{Pr}\left[J_u \geq \gamma^3\Delta/10\right] \leq \exp\left(-\gamma^3\Delta/10\right).$$

Let L_u be the number of vertices in $z \in N(u)$ which are occupied in X_0. Since we have that X_0 is $(400, R)$-nice at w for radius $R \gg 100$ and $u \in \mathcal{B}_{100}(w)$, it holds that $L_u \leq 400\Delta/\log\Delta$. Since $|A_u| = J_u + L_u$ we get that $\mathbf{Pr}\left[|A_u| \geq \gamma^3\Delta\right] \leq \exp(-\gamma^3\Delta/10)$. Taking a union bound over the at most Δ^{100} vertices in $B_{100}(w)$ we get that

$$\mathbf{Pr}\left[\exists u \in B_{100}(w) \ s.t.\ |A_u| \geq \gamma^3\Delta\right] \leq \Delta^{100}\exp\left(-\gamma^3\Delta/10\right) \leq \exp\left(-\gamma^3\Delta/20\right),$$

where the last inequality holds for sufficiently large Δ. Working in the same way we get that

$$\mathbf{Pr}\left[\exists u \in B_{100}(w) \ s.t.\ |A_u^*| \geq \gamma^3\Delta\right] \leq \exp\left(-\gamma^3\Delta/20\right).$$

Combining the two inequalities above, there exists $C_7 = C_7(\gamma) > 0$ such that

$$\mathbf{Pr}\left[\mathcal{B}_4\right] \leq \exp\left(-\Delta/C_7\right). \tag{112}$$

Plugging (112), (111), (109), (108), and (107) into (106), we get that

$$\mathbf{Pr}\left[\mathcal{A}\right] \leq \exp\left(-\Delta/C_8\right) \tag{113}$$

for appropriate $C_8 > 0$. The lemma follows by plugging (113) and (111) into (105). ∎

10. Proof of local uniformity—Proof of Theorems 11 and 18. In light of Lemma 21, Theorem 11 follows as a corollary from Theorem 18. For this reason we focus on proving Theorem 18. We will use Lemmas 21 and 29 to complete the proof of Theorem 18.

For an independent set σ of G and $w \in V$, recall that $\mathbf{R}(\sigma, w) = \prod_{z \in N(w)}(1 - \frac{\lambda}{1+\lambda}\mathbf{U}_{z,v}(\sigma))$, where $\mathbf{U}_{z,w}(\sigma) = \mathbf{1}\{\sigma \cap (N(z) \setminus \{w\}) = \emptyset\}$.

LEMMA 32. *Let $\epsilon > 0$, R, C, and λ be as in Theorem* 18. *Let (X_t) be the continuous time Glauber dynamics on the hard-core model with fugacity λ and underlying graphs G. If X_0 is $(400, R)$-nice at $w \in V$, then we have that*

$$\mathbf{Pr}\left[(\forall t \in \mathcal{I}) \quad |\mathbf{R}(X_t, w) - \omega^*(w)| \le \epsilon/10\right] \ge 1 - \exp\left(-20\Delta/C\right), \tag{114}$$

where $\mathcal{I} = [Cn, n\exp(\Delta/C)]$.

The proof of Lemma 32 makes use of the following result, which is the Glauber dynamics version of Lemma 12 in section 4.1.

LEMMA 33. *For $\delta, \gamma > 0$, let $\Delta_0 = \Delta_0(\delta, \gamma), C = C(\delta, \gamma)$, $\hat{C} = \hat{C}(\delta, \gamma)$. For all graphs $G = (V, E)$ of maximum degree $\Delta \ge \Delta_0$ and girth ≥ 7, all $\lambda < (1-\delta)\lambda_c(\Delta)$, let (X_t) be the continuous time Glauber dynamics on the hard-core model.*

Let X_0 be $(400, R)$-nice at w for radius $R \le \Delta^{9/10}$. Then, for $x \in B_{R/2}(w)$ and $I = [t_0, t_1]$, where $t_0 = Cn$, it holds that

$$\mathbf{Pr}\left[(\forall t \in I) \quad \left|\mathbf{R}(X_t, x) - \exp\left(-\frac{\lambda}{1+\lambda}\sum_{z \in N(x)} \mathbb{E}_{t_z}\left[\mathbf{R}(X_{t_z}, z)\right]\right)\right| \le \gamma\right]$$
$$\ge 1 - \left(1 + \frac{t_1 - t_0}{n}\right)\exp\left(-\Delta/\hat{C}\right),$$

where $\mathbb{E}_{t_z}[\mathbf{R}(X_{t_z}, z)]$ is the expectation w.r.t. random time t_z, the last time that vertex z is updated prior to time t.

Note that $\mathbb{E}_{t_z}[\mathbf{R}(X_{t_z}, z)] = \exp(-t/n)\ \mathbf{R}(X_0, z) + \int_0^t \mathbf{R}(X_s, z) n \exp((s-t)/n) ds$. The proof of Lemma 33 is lengthy for this reason we present it separately in section 10.1.

Proof of Lemma 32. Recall that $\mathcal{I} = [Cn, n\exp(\Delta/C]$. Let $R = \lfloor 30\delta^{-1}\log(6\epsilon^{-1})\rfloor$. Assume that C is sufficiently large such that $C = (R+1)C_1$, where C_1 is specified later. Let $T_0 = (R+1)C_1 n$ and $T_1 = \exp(\Delta/C)$. Finally, for $i \le R$ let $\mathcal{I}_i := [T_0 - iC_1 n, T_1]$.

Consider the continuous time Glauber dynamics (X_t). Also, consider times $t \ge T_0 - RC_1 n$. For each such time t and positive integer $i \le R$, we define

$$\alpha_i := \max\left|\Psi(\mathbf{R}(X_t, x)) - \Psi(\omega^*(x))\right|,$$

where Ψ is defined in (5). The maximum is taken over all $t \in \mathcal{I}_i$ and over all vertices $x \in B_i(w)$.

An elementary observation about α_i is that $\alpha_i \le 3$ for every $i \le R$. To see why this holds, note the following: For any $z \in V$ and any independent sets σ, it holds that

$$\mathbf{R}(\sigma, z) = \prod_{r \in N(z)}\left(1 - \frac{\lambda \cdot \mathbf{U}_{r,z}(\sigma)}{1+\lambda}\right) \ge (1+\lambda)^{-\Delta} \ge e^{-\lambda\Delta} \ge \mathrm{e}^{-\mathrm{e}},$$

where in the last inequality we use the fact that Δ is sufficiently large, i.e., $\Delta > \Delta_0(\epsilon, \delta)$ and $\lambda < e/\Delta$. Furthermore, using the same arguments as above we get that

$\omega^*(z) \geq e^{-e}$, as well. Since for any $x \in V$ and any independent sets σ, we have $\mathbf{R}(\sigma, x), \omega^*(x) \in [e^{-e}, 1]$, (6) implies $\alpha_i \leq C_0 = 3$ for every $i \leq R$.

We prove our result by showing that typically α_0 is very small. Then, the lemma follows by using standard arguments. We use an inductive argument to show that α_0 is very small. We start by using the fact that $\alpha_R \leq C_0$. Then we show that with sufficiently large probability, if $\alpha_{i+1} \geq \epsilon/20$, then $\alpha_i \leq (1-\gamma)\alpha_{i+1}$, where $0 < \gamma < 1$.

For any $i \leq R$, we use the fact that there exists $\hat{C} > 0$ such that with probability at least $1 - \exp(-\Delta/\hat{C})$ the following is true: For every vertex $z \in B_i(w)$ it holds that

$$(\forall t \in \mathcal{I}_i) \quad \left| \mathbf{R}(X_t, z) - \exp\left(-\frac{\lambda}{1+\lambda} \sum_{r \in N(z)} \tilde{\omega}(r) \right) \right| < \frac{\epsilon^2 \delta}{40}, \tag{115}$$

where

$$\tilde{\omega}_t(r) = \exp(-C_1) \cdot \mathbf{R}(X_{t-C_1 n}, r) + \int_{t-C_1 n}^{t} \mathbf{R}(X_s, r) n \exp\left[(s - C_1 n)/n\right] ds. \tag{116}$$

Equation (115) is implied by Lemma 33.

Fix some $i \leq R$, $z \in B_i(w)$ and time $s \in \mathcal{I}_i$. We consider X_s by conditioning on $X_{s-C_1 n}$. From the definition of the quantity α_{i+1} we get the following: For any $x \in B_{i+1}(w)$ consider the quantity $\tilde{\omega}_s(x)$. We have that

$$D_{v,i+1}(\tilde{\omega}_s, \omega^*) \leq \alpha_{i+1}. \tag{117}$$

We will show that if (115) holds for $\mathbf{R}(X_s, z)$, where $z \in B_i(w)$, and $\alpha_{i+1} \geq \epsilon/20$, then we have that

$$|\Psi(\mathbf{R}(X_s, z)) - \Psi(\omega^*(z))| \leq (1 - \delta/24)\alpha_{i+1}.$$

For proving the above inequality, first note that if $\mathbf{R}(X_s, z)$ satisfies (115), then (6) implies that

$$\left| \Psi(\mathbf{R}(X_s, z)) - \Psi\left(\exp\left(-\frac{\lambda}{1+\lambda} \sum_{r \in N(z)} \tilde{\omega}_s(r) \right) \right) \right| \leq \frac{\delta \epsilon^2}{12}. \tag{118}$$

Furthermore, we have that

$$\begin{aligned}
&|\Psi(\mathbf{R}(X_s, z)) - \Psi(\omega^*(z))| \\
&\quad \leq \frac{\delta\epsilon^2}{12} + \left| \Psi\left(\exp\left(-\frac{\lambda}{1+\lambda} \sum_{r \in N_z} \tilde{\omega}_s(r) \right) \right) - \Psi(\omega^*(z)) \right| \qquad \text{(from (118))} \\
&\quad \leq \frac{\delta\epsilon^2}{12} + \left| \Psi\left(\prod_{r \in N(z)} \left(1 - \frac{\lambda \tilde{\omega}_s(r)}{1+\lambda} \right) \right) - \Psi(\omega^*(z)) \right| \\
&\qquad + \left| \Psi\left(\prod_{r \in N(z)} \left(1 - \frac{\lambda \tilde{\omega}_s(r)}{1+\lambda} \right) \right) - \Psi\left(\exp\left(-\frac{\lambda}{1+\lambda} \sum_{r \in N(z)} \tilde{\omega}(r) \right) \right) \right|, \qquad (119)
\end{aligned}$$

where the last derivation follows from the triangle inequality.

From our assumptions about λ, Δ and the fact that $\tilde{\omega}_s(r) \in [e^{-e}, 1]$ for $r \in N(z)$, we have that

$$\left| \prod_{r \in N(z)} \left(1 - \frac{\lambda \tilde{\omega}_s(r)}{1+\lambda}\right) - \exp\left(-\lambda \sum_{r \in N(z)} \frac{\tilde{\omega}_s(r)}{1+\lambda}\right) \right| \le \frac{10}{\Delta}.$$

The above inequality and (6) imply that

$$\left| \Psi\left(\prod_{r \in N(z)} \left(1 - \frac{\lambda \tilde{\omega}_s(r)}{1+\lambda}\right)\right) - \Psi\left(\exp\left(-\frac{\lambda}{1+\lambda} \sum_{r \in N(z)} \tilde{\omega}_s(r)\right)\right) \right| \le \frac{30}{\Delta}.$$

Plugging the inequality above into (119) we get that

$$\begin{aligned} |\Psi(\mathbf{R}(X_s, z)) - \Psi(\omega^*(z))| &\le \frac{\delta\epsilon^2}{12} + \frac{30}{\Delta} + \left| \Psi\left(\prod_{r \in N(z)} \left(1 - \frac{\lambda \tilde{\omega}_s(r)}{1+\lambda}\right)\right) - \Psi(\omega^*(z)) \right| \\ &\le \frac{\delta\epsilon^2}{12} + \frac{30}{\Delta} + \left| \Psi\left(\prod_{r \in N(z)} \left(\frac{1}{1+\lambda \tilde{\omega}_s(r)}\right)\right) - \Psi(\omega^*(z)) \right| \\ &\quad + 3\left| \prod_{r \in N(z)} \left(1 - \frac{\lambda \tilde{\omega}_s(r)}{1+\lambda}\right) + \prod_{r \in N(z)} \left(\frac{1}{1+\lambda \tilde{\omega}_s(r)}\right) \right| \qquad (120) \\ &\le \frac{\delta\epsilon^2}{12} + \frac{60}{\Delta} + D_{v,i}(F(\tilde{\omega}), \omega^*), \qquad (121) \end{aligned}$$

where we derive (120) by applying the triangle inequality and (6). Equation (121) follows by noting that for any $r \in N(z)$ we have $(\frac{1}{1+\lambda\tilde{\omega}_s(r)}) - (1 - \frac{\lambda\tilde{\omega}_s(r)}{1+\lambda}) \le (e/\Delta)^2$, $|N(z)| \le \Delta$, and Δ is large. Finally, in (38) the function F is defined in (3). Since $\tilde{\omega}_s$ satisfies (117), Lemma 5 implies that

$$D_{v,i}(F(\tilde{\omega}), \omega^*) \le (1 - \delta/6)\alpha_{i+1}. \qquad (122)$$

Plugging (122) into (121) we get that

$$|\Psi(\mathbf{R}(X_s, z)) - \Psi(\omega^*(z))| \le \frac{\delta\epsilon^2}{12} + \frac{60}{\Delta} + (1-\delta/6)\alpha_{i+1} \le (1 - \delta/24)\alpha_{i+1}, \qquad (123)$$

where the last inequality follows if we have $\alpha_{i+1} \ge \epsilon/20$ and Δ sufficiently large. Note that (123) holds provided that $\mathbf{R}(X_s, z)$ satisfies (115).

To bound α_i we have to take the maximum over all times $t \in \mathcal{I}_i$ and vertices $z \in B_i(w)$. So far, i.e., in (123), we only considered a fixed time $s \in \mathcal{I}_i$ and a fixed vertex z.

Consider, now, a partition of $\mathcal{I}_i$ into subintervals each of length $\frac{\epsilon^4\eta}{200\Delta}n$, where the last part can be of smaller length. Let $T(j)$ be the jth part for $j \in \{1, \ldots, \lceil 200C_1 \Delta/(\epsilon^4\eta)\rceil\}$. For some some vertex $x \in V$, each $r \in N(x)$ is updated during the time period $T(j)$ with probability less than $\frac{\epsilon^4\eta}{100\Delta}$, independently of the other vertices.

Chernoff's bounds imply that with probability at least $1 - \exp(-\Delta\epsilon^3/3)$, the number of vertices in $S_2(x)$ which are updated during the interval $T(j)$ is at most $\Delta\epsilon^3/3$. Furthermore, changing any $\Delta\epsilon^2/3$ variables in $S_2(x)$ can only change $\mathbf{R}(X_s, x)$

by at most $\epsilon^2/3$. Consequently, $\Psi(\mathbf{R}(X_s, x))$ can change by only ϵ^2 within $T(j)$. From a union bound over all subintervals $T(j)$ and all vertices $x \in B_i(w)$, there exists sufficiently large $C > 0$ such that

$$\mathbf{Pr}\left[\alpha_i = \max\left\{3\epsilon^2 + (1-\delta/24)\alpha_{i+1}, \epsilon/20\right\}\right] \geq 1 - \exp\left(-52\Delta/C\right).$$

The fact that $\alpha_R \leq C_0$ and $R = \lfloor 20\delta^{-1}\log(6\epsilon^{-1})\rfloor$ implies the following: With probability at least $1 - \exp(-50\Delta/C)$ for every $t \in \mathcal{I}$ it holds that $\alpha_0 \leq \epsilon/30$. In turn, (6) implies that

$$|\mathbf{R}(X_t, v) - \omega^*(v)| \leq \epsilon/11. \tag{124}$$

The lemma follows. □

We conclude the technical results for Theorem 18 by proving the following lemma.

LEMMA 34. *Let $\epsilon > 0$, R, $\mathcal{I}$, and λ be as in Theorem 18. Let (X_t) be the continuous time Glauber dynamics on the hard-core model with fugacity λ and underlying graphs G. Assume that X_0 is $(400, R)$-nice at v. Then, for any $t \in \mathcal{I}$, any $\gamma > 0$, there is $\hat{C} = \hat{C}(\gamma) > 0$ such that*

$$\mathbf{Pr}\left[\left|\mathbf{W}(X_t, v) - \sum_{z\in N(v)} \Phi(z)\cdot\mathbb{E}_{t_z}\left[\mathbf{R}(X_{t_z}, z)\right]\right| > \gamma\Delta\right] < \exp\left(-\Delta/\hat{C}\right).$$

Recall that $\mathbb{E}_{t_z}[\mathbf{R}(X_{t_z}, z)]$ is the expectation w.r.t. t_z the time when z was last updated prior to time t, i.e., $\mathbb{E}_{t_r}[\mathbf{R}(X_{t_z}, z)] = \exp(-t/n)\mathbf{R}(X_0, z) + \int_0^t \mathbf{R}(X_s, z) n \exp[(s-t)/n] ds$.

Proof. Consider, first, the graph G_v^* and the dynamics (X_t^*) such that $X_0^* = X_0$. Condition on X_0^* and on X_t^* restricted to $V \setminus B_2(x)$ for all $t \in \mathcal{I}$. Denote this conditional information by $\mathcal{F}$.

First we are going to show that $\mathbb{E}\left[\mathbf{W}(X_t^*, v) \mid \mathcal{F}\right]$ and $\sum_{z\in N(v)} \Phi(z)\cdot\mathbb{E}\left[\mathbf{R}(X_t^*, z) \mid \mathcal{F}\right]$ are very close. From the definition of $\mathbf{W}(X_t^*, v)$ we have that

$$\mathbb{E}\left[W(X_t^*, v) \mid \mathcal{F}\right] = \sum_{z\in N(v)} \Phi(z)\cdot\mathbb{E}\left[\mathbf{U}_{z,v}(X_t^*) \mid \mathcal{F}\right].$$

Let $c > 0$ be such that $t/n = c$. For $\zeta > 0$ whose value is going to be specified later, let $H(v) \subseteq N(v)$ be such that $z \in H(v)$ is $|N(z) \cap X_0^*| \geq \zeta^{-1}$. In (148) and (149) we have shown that for $z \notin H(v)$ it holds that

$$\left|\mathbb{E}\left[\mathbf{U}_{z,v}(X_t^*) \mid \mathcal{F}\right] - \mathbb{E}_{t_z}\left[\mathbf{R}(X_{t_z}^*, z) \mid \mathcal{F}\right]\right| \leq \theta, \tag{125}$$

where $0 < \theta = \theta(c, \zeta) < 20(\zeta e^c)^{-1}$ while (as in we previously defined)

$$\mathbb{E}_{t_z}\left[\mathbf{R}(X_{t_z}^*, z) \mid \mathcal{F}\right] = \exp(-t/n)\mathbf{R}(X_0^*, z) + \int_0^t \mathbf{R}(X_s^*, z) n \exp\left[(s-t)/n\right] ds.$$

Since X^*_0 is $(400,R)$-nice at v it holds that $|H(v)| \le 400\zeta\varDelta$. We have that

$$\begin{aligned}
&\left|\mathbb{E}\left[\mathbf{W}(X^*_t,v)\mid\mathcal{F}\right]-\sum_{z\in N(v)}\varPhi(z)\cdot\mathbb{E}_{t_z}\left[\mathbf{R}(X^*_{t_z},z)\mid\mathcal{F}\right]\right|\\
&\quad\le\left|\mathbb{E}\left[\mathbf{W}(X^*_t,v)\mid\mathcal{F}\right]-\sum_{z\notin H(v)}\varPhi(z)\cdot\mathbb{E}_{t_z}\left[\mathbf{R}(X^*_{t_z},z)\mid\mathcal{F}\right]\right|\\
&\qquad+\sum_{z\in H(v)}\varPhi(z)\cdot\mathbb{E}_{t_z}\left[\mathbf{R}(X^*_{t_z},z)\mid\mathcal{F}\right]\\
&\quad\le\left|\mathbb{E}\left[\mathbf{W}(X^*_t,v)\mid\mathcal{F}\right]-\sum_{z\notin H(v)}\varPhi(z)\cdot\mathbb{E}_{t_z}\left[\mathbf{R}(X^*_{t_z},z)\mid\mathcal{F}\right]\right|\\
&\qquad+5000\zeta\varDelta\qquad(\text{since }\max_z\varPhi(z)\le 12)\\
&\quad\le(12\theta+5000\zeta)\,\varDelta\qquad\qquad(\text{from (125)}).
\end{aligned}\tag{126}$$

The fact that $\max_z \varPhi(z) \le 12$ is from Theorem 9.

We proceed by showing that $W(X^*_t,v)$ is sufficiently well concentrated about its expectation. Conditioning on $\mathcal{F}$ the random variables $\mathbf{U}_{z,v}(X^*_t)$ for $z\in N(v)$ are fully independent. From Chernoff's bounds, there exists appropriate $C_1>0$ such that

$$\mathbf{Pr}\left[\left|\mathbf{W}(X^*_t,v)-\mathbb{E}\left[\mathbf{W}(X^*_t,v)\mid\mathcal{F}\right]\right|>\gamma\varDelta/100\right]\le\exp\left(-\varDelta/C_1\right).\tag{127}$$

From (127) and (126) there exists $C_2>0$ such that that

$$\mathbf{Pr}\left[\left|W(X^*_t,v)-\sum_{z\in N(v)}\varPhi(z)\cdot\mathbb{E}_{t_z}\left[\mathbf{R}(X^*_{t_z},z)\mid\mathcal{F}\right]\right|\ge\gamma\varDelta/50\right]\le\exp\left(-\varDelta/C_2\right).\tag{128}$$

Furthermore, using Lemma 29 with error parameter γ^2, i.e., $|(X^*_t\oplus X_t)\cap B_2(v)|\le\gamma^2\varDelta$, we get the following: There exists appropriate $C_3=C_3(\gamma)>0$ such that

$$\mathbf{Pr}\left[\left|\mathbf{W}(X^*_t,v)-\mathbf{W}(X_t,v)\right|\le\gamma\varDelta/40\right]\ge 1-\exp\left(-\varDelta/C_3\right).\tag{129}$$

Also, (from Lemma 29 again) with probability at least $1-\exp\left(-\varDelta/C_3\right)$ it holds that

$$\left|\int_0^t\mathbf{R}(X_s,z)n\exp\left[(s-t)/n\right]ds-\int_0^t\mathbf{R}(X^*_s,z)n\exp\left[(s-t)/n\right]ds\right|\le\gamma/600\tag{130}$$

for every $z\in N(v)$. The above follows by using the fact that changing the spin of any $\gamma^2\varDelta$ vertices in $X^*_t(B_2(z))$ changes $\mathbf{R}(X^*_t,z)$ by at most $\gamma/1000$.

Noting that $\varPhi(z)\le 12$, for any z, the lemma follows by combining (130), (129), and (128). ∎

Using Lemmas 32 and 34, in this section we prove Theorem 18. Recall that Theorem 11 follows as a corollary of Theorem 18 and Lemma 21.

Proof of Theorem 18. For a vertex $u\in N(v)$ consider G^*_u. Consider also the continuous time dynamics (X^*_t) such that $X^*_0=X_0$.

We condition on the restriction of (X_t^*) to $V \setminus B_2(u)$ for every $t \in \mathcal{I}$. We denote this by $\mathcal{F}$. Fix some $t \in \mathcal{I}$. Since $u \in B_R(v)$ and X_0 is $(400, R)$-nice at v, we get that

$$\begin{aligned}
&\mathbb{E}_s\left[\mathbf{R}(X_s^*, u) \mid \mathcal{F}\right] \\
&\quad = \exp(-t/n)\mathbf{R}(X_0^*, u) + \int_0^t \mathbf{R}(X_t^*, u) n \exp\left((s-t)/n\right) \\
&\quad = \mathbb{E}_s\left[\exp\left(-\frac{\lambda}{1+\lambda}\sum_{z\in N(u)} \mathbf{U}_{z,u}(X_s^*) + O\left(1/\Delta\right)\right) \mid \mathcal{F}\right] \\
&\quad = \exp\left(-\frac{\lambda}{1+\lambda}\sum_{z\in N(u)} \mathbb{E}_s\left[\mathbf{U}_{z,u}(X_s^*) \mid \mathcal{F}\right] + O\left(1/\Delta\right)\right) \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{(due to conditioning on } \mathcal{F}) \\
&\quad \leq \exp\left(-\frac{\lambda}{1+\lambda}\sum_{z\in N(u)} \mathbb{E}_s\left[\mathbf{R}(X_s^*, z) \mid \mathcal{F}\right] + \theta\lambda\Delta + O\left(1/\Delta\right)\right),
\end{aligned} \tag{131}$$

where in the last inequality we use (125). Note that so as apply (125) $X_0^*(u)$ should be sufficiently "light." This is guaranteed from our assumption that $u \in B_R(v)$ and X_0 is $(400, R)$-nice at v.

Furthermore, (115) and Lemma 29 imply the following: There exists $C_1 > 0$ such that with probability at least $1 - \exp(-\Delta/C_1)$, we have that

$$(\forall t \in \mathcal{I}) \quad \left|\mathbf{R}(X_t^*, u) - \exp\left(-\frac{\lambda}{1+\lambda}\sum_{r\in N^*(u)} \hat{\omega}(r)\right)\right| < \gamma, \tag{132}$$

where

$$\hat{\omega}(r) = \exp(-t/n)\mathbf{R}(X_0^*, r) + \int_0^t \mathbf{R}(X_s^*, r) n \exp\left[(s-t)/n\right] ds.$$

Note that for every $r \in N^*(u)$ we have $\hat{\omega}(r) = \mathbb{E}_{t_r}[\mathbf{R}(X_{t_r}^*, r)|\mathcal{F}]$. Using this observation, we plug (131) into (132) and get

$$\mathbf{Pr}\left[\left|\mathbf{R}(X_t^*, u) - \hat{\omega}(u)\right| \geq 10e\theta + \gamma\right] \leq \exp\left(-\Delta/C_1\right). \tag{133}$$

In the above inequality we used the fact that $\lambda\Delta < 2e$.

Consider the continuous time Glauber dynamics (X_t). From Lemma 29 and (133) there exists $C_3 > 0$ such that for X_t the following holds:

$$\mathbf{Pr}\left[\left|\mathbf{R}(X_t, u) - \tilde{\omega}(u)\right| \geq 20e\theta + 2\gamma\right] \leq \exp\left(-\Delta/C_3\right), \tag{134}$$

where

$$\tilde{\omega}(z) = \exp(-t/n)\mathbf{R}(X_0, z) + \int_0^t \mathbf{R}(X_s, z) n \exp\left[(s-t)/n\right] ds.$$

Furthermore a simple union bound over $u \in N(v)$ and (134) gives that

$$\mathbf{Pr}\left[\forall u \in N(v) \quad \left|\mathbf{R}(X_t, u) - \tilde{\omega}(u)\right| \geq 20e\theta + 2\gamma\right] \leq \Delta \exp\left(-\Delta/C_3\right). \tag{135}$$

Taking sufficiently small θ, γ in (135) and using Lemma 34 we get that

$$\mathbf{Pr}\left[\left|\mathbf{W}(X_t, v) - \sum_{w\in N(v)} \Phi(z)\cdot\mathbf{R}(X_t, w)\right| > \epsilon\Delta/15\right] \leq \exp\left(-\Delta/C_4\right) \tag{136}$$

for appropriate $C_4 > 0$. Furthermore, applying Lemma 32 for each $w \in N(v)$ and using (136) yields

$$\mathbf{Pr}\left[\left|\mathbf{W}(X_t, v) - \sum_{w\in N(v)} \Phi(w)\cdot\omega^*(w)\right| > \epsilon\Delta/2\right] \le \exp\left(-\Delta/C_5\right) \tag{137}$$

for appropriate $C_5 > 0$. The above inequality establishes the desired result for a fixed $t \in \mathcal{I}$.

Now we will prove that (137) holds for all $t \in \mathcal{I}$. Consider a partition of the time interval $\mathcal{I}$ into subintervals each of length $\frac{\psi^2}{\Delta}n$, where the last part can be of smaller length. The quantity $\psi > 0$ is going to be specified later. Also, let $T(j)$ be the jth part.

Each $z \in B_2(v)$ is updated at least once during the time period $T(j)$ with probability less than $2\frac{\psi^2}{\Delta}$, independently of the other vertices. Note that $|B_2(v)| \le \Delta^2$. Clearly, the number of vertices in $B_2(v)$ which are updated during $T_i(j)$ is dominated by $\mathcal{B}(\Delta^2, 2\psi^2/\Delta)$. Chernoff's bounds imply that with probability at least $1 - \exp(-20\Delta\psi^2)$, the number of vertices in $B_2(v)$ which are updated during the interval $T(j)$ is at most $20\psi^2\Delta$. Furthermore, changing any $2\Delta\psi^2$ variables in $B_2(v)$ can only change the weighted sum of unblocked vertices in N_v by at most $20C_0\psi^2\Delta$. Taking sufficiently small $\psi > 0$ we get the following:

$$\mathbf{Pr}\left[\left|\mathbf{W}(X_t, v) - \sum_{w\in N(v)} \Phi(w)\cdot\omega^*(w)\right| > \epsilon\Delta\right] \le \exp\left(-2\Delta/C_b\right). \tag{138}$$

The above completes the proof of Theorem 18 for the case where (X_t) is the continuous time process.

The discrete time result follows by working as follows: instead of $\mathbf{W}(X_t, v)$ we consider the "normalized" variable $\Lambda(X_t, v) = \frac{\mathbf{W}(X_t,v)}{\Delta}$. Rephrasing (137) in terms of $\Lambda(X_t, v)$ we have, for a specific $t \in \mathcal{I}$,

$$\mathbf{Pr}\left[\left|\Lambda(X_t, v) - \Delta^{-1}\sum_{w\in N(v)} \Phi(w)\cdot\omega^*(w)\right| > \epsilon/2\right] \le \exp\left(-\Delta/C_5\right). \tag{139}$$

Note that $\Lambda(X_t, v)$ satisfies the Lipschitz and total influence conditions of Lemma 28. Hence by Lemma 28 the result for the discrete time process holds. ∎

10.1. Approximate recurrence for Glauber dynamics—Proof of Lemma 33. Consider G^*_x and let (X^*_t) be the Glauber dynamics on G^*_x with fugacity $\lambda > 0$ and let $X^*_0 = X_0$. Also assume that (X^*_t) and (X_t) are maximally coupled.

Condition on X^*_0, and let $\mathcal{F}$ be the σ-algebra generated by X^*_t restricted to $V \setminus B_2(x)$ for all $t \in I$. Fix $t \in I$. Let $c > 0$ be such that $t/n = c$, i.e., c is a large constant. Recalling the definition of $\mathbf{R}(X^*_t, x)$, we have that

$$\begin{aligned}\mathbf{R}(X^*_t, x) &= \prod_{z\in N(x)}\left(1 - \frac{\lambda}{1+\lambda}\mathbf{U}_{z,x}(X^*_t)\right)\\ &= \exp\left(-\frac{\lambda}{1+\lambda}\sum_{z\in N(x)}\mathbf{U}_{z,x}(X^*_t) + O\left(1/\Delta\right)\right). \qquad (140)\end{aligned}$$

Let $\mathbf{Q}(X^*_t) = \sum_{z\in N(x)}\mathbf{U}_{z,x}(X^*_t)$. Conditional on $\mathcal{F}$, the quantity $\mathbf{Q}(X^*_t)$ is a sum of $|N(x)|$ many independent random variables in $[0,1]$. Applying Azuma's inequality, for $0 \le \gamma \le (3e)^{-1}$, we have

$$\mathbf{Pr}\left[\left|\mathbb{E}\left[\mathbf{Q}(X_t^*) \mid \mathcal{F}\right] - \mathbf{Q}(X_t^*)\right| \ge \gamma\Delta\right] \le 2\exp\left(-\gamma^2\Delta/2\right). \tag{141}$$

Combining the fact that $\mathbb{E}\left[\mathbf{Q}(X_t^*) \mid \mathcal{F}\right] = \sum_{z\in N(x)} \mathbb{E}\left[\mathbf{U}_{z,x}(X_t^*) \mid \mathcal{F}\right]$ with (141) and (140) we get that

$$\mathbf{Pr}\left[\left|\mathbf{R}(X_t^*, x) - \exp\left(-\frac{\lambda}{1+\lambda}\sum_{z\in N(x)} \mathbb{E}\left[\mathbf{U}_{z,x}(X_t^*) \mid \mathcal{F}\right]\right)\right| \ge 3\gamma\lambda\Delta\right] \le 2\exp\left(-\gamma^2\Delta/2\right). \tag{142}$$

For every $z \in N^*(x)$, it holds that

$$\begin{aligned}
&\mathbb{E}\left[\mathbf{U}_{z,x}(X_t^*) \mid \mathcal{F}\right]\\
&\quad = \prod_{y\sim N(z)\setminus\{x\}} \mathbb{E}\left[\mathbf{1}\{y \notin X_t^*\} \mid \mathcal{F}\right]\\
&\quad = \prod_{y\sim N(z)\setminus\{x\}} \left(\Pr[t_y = 0]\cdot \mathbf{1}\{y\notin X_0^*\} + \mathbb{E}\left[\mathbf{1}\{y\notin X_t^*\}\cdot\mathbf{1}\{t_y>0\} \mid \mathcal{F}\right]\right),
\end{aligned} \tag{143}$$

where t_y is the time that vertex y is last updated prior to time t and it is defined to be equal to zero if y is not updated prior to t. Note, for any $0 \le s \le t$, it holds that $\Pr[t_y \le s] = e^{-(t-s)/n}$. Also, we have that

$$\begin{aligned}
\mathbb{E}\left[\mathbf{1}\{y\notin X_t^*\}\cdot\mathbf{1}\{t_y>0\} \mid \mathcal{F}\right] &= \mathbb{E}\left[\mathbb{E}\left[\mathbf{1}\{y\notin X_t^*\}\cdot\mathbf{1}\{t_y>0\} \mid \mathcal{F}, t_y\right] \mid \mathcal{F}\right]\\
&= \int_0^t \left(1 - \frac{\lambda}{1+\lambda}\mathbf{U}_{y,z}(X_s^*)\right) n\exp\left[(s-t)/n\right] ds,
\end{aligned} \tag{144}$$

where the last equality follows because we are using G_x^* and (X_t^*). The use of G^* and (X_t^*) ensures that the configuration in $V \setminus B_2(x)$ is never affected by that in $B_2(x)$. For this reason, if y is updated at time $s \in I$, then the probability for it to be occupied, given $\mathcal{F}$, is exactly $\frac{\lambda}{(1+\lambda)}\mathbf{U}_{y,z}(X_s^*)$. That is, the configuration outside $B_2(x)$ does not provide any information for y but the value of $\mathbf{U}_{y,z}(X_s^*)$.

Plugging (144) into (143) we get that

$$\begin{aligned}
&\mathbb{E}\left[\mathbf{U}_{z,x}(X_t^*) \mid \mathcal{F}\right]\\
&= \prod_{y\sim N(z)\setminus\{x\}} \left[\exp(-t/n)\,\mathbf{1}\{y\notin X_0^*\} - \int_0^t \left(1-\frac{\lambda}{1+\lambda}\mathbf{U}_{y,z}(X_s^*)\right) n\exp\left[(s-t)/n\right] ds\right]\\
&= \prod_{y\sim N(z)\setminus\{x\}} \left[1 - \exp(-t/n)\,\mathbf{1}\{y\in X_0^*\} - \int_0^t \frac{\lambda}{1+\lambda}\mathbf{U}_{y,z}(X_s^*) n\exp\left[(s-t)/n\right] ds\right].
\end{aligned} \tag{145}$$

For appropriate $\zeta \in (0,1)$, which we define later, let $H(x) \subseteq N^*(x)$ be such that $z \in H(x)$ if $|N^*(z) \cap X_0^*| \ge 1/\zeta$.

Noting that each integral in (145) is less than λ, for every $z \notin H(x)$, we get that

$$\mathbb{E}\left[\mathbf{U}_{z,x}(X_t^*) \mid \mathcal{F}\right] = (1+\delta)\prod_{y\in N(z)\setminus\{x\}} \left(1 - \int_0^t \frac{\lambda}{1+\lambda}\mathbf{U}_{y,z}(X_s^*) n\exp\left[(s-t)/n\right] ds\right), \tag{146}$$

where $|\delta| \le 4(\zeta e^c)^{-1}$.

Recall that for some vertex y in G_x^* we let $\mathbb{E}_{t_y}[\cdot \mid \mathcal{F}]$ denote the expectation w.r.t. t_y, the random time that y is updated prior to time t. It holds that

$$\mathbb{E}_{t_y}[\mathbf{U}_{y,z}(X^*_{t_y})|\mathcal{F}] = \exp(-t/n)\mathbf{U}_{y,z}(X^*_0) + \int_0^t \mathbf{U}_{y,z}(X^*_s) n \exp\left[(s-t)/n\right] ds.$$

For every $y \in N(z)\backslash\{x\}$, where $z \notin H(x)$ it holds that

$$\begin{aligned}\mathbb{E}_{t_y}\left[\mathbf{U}_{y,z}(X^*_{t_y}) \mid \mathcal{F}\right] - \int_0^t \mathbf{U}_{y,z}(X^*_s) n \exp\left[(s-t)/n\right] ds &= \exp\left(-t/n\right)\mathbf{U}_{y,z}(X^*_0) \\ &\le \exp\left(-t/n\right) \le \exp(-c).\end{aligned} \tag{147}$$

Since $\lambda < e/\Delta$, (146) implies that there is a quantity θ with $0 < \theta \le 20(\zeta e^c)^{-1}$, such that

$$\begin{aligned}\mathbb{E}\left[\mathbf{U}_{z,x}(X^*_t) \mid \mathcal{F}\right] &\le \prod_{y \in N(z)\backslash\{x\}} \left(1 - \int_0^t \frac{\lambda}{1+\lambda}\mathbf{U}_{y,z}(X^*_s) n \exp\left[(s-t)/n\right] ds\right) + \theta/2 \\ &\le \prod_{y \in N(z)\backslash\{x\}} \left(1 - \mathbb{E}_{t_y}\left[\frac{\lambda}{1+\lambda}\mathbf{U}_{y,z}(X^*_{t_y}) \mid \mathcal{F}\right]\right) + \theta \qquad \text{(from (147))} \\ &= \prod_{y \in N(z)\backslash\{x\}} \left(1 - \mathbb{E}_{s}\left[\frac{\lambda}{1+\lambda}\mathbf{U}_{y,z}(X^*_{s}) \mid \mathcal{F}\right]\right) + \theta,\end{aligned}$$

where in the last derivation, we substituted the variables t_y for $y \in N(z) \setminus \{x\}$ with a new random variable s which follows the same distribution as t_y. Note that the variables t_y are identically distributed.

Given the σ-algebra $\mathcal{F}$, the variables $\mathbf{U}_{y,z}(X^*_s)$ for $y \in N(z)\backslash\{x\}$ are independent with each other; this yields

$$\begin{aligned}\mathbb{E}\left[\mathbf{U}_{z,x}(X^*_t) \mid \mathcal{F}\right] &= \mathbb{E}_s\left[\prod_y \left(1 - \frac{\lambda}{1+\lambda}\mathbf{U}_{y,z}(X^*_s)\right) \mid \mathcal{F}\right] + \theta \\ &= \mathbb{E}_s\left[\mathbf{R}(X^*_s, z) \mid \mathcal{F}\right] + \theta,\end{aligned} \tag{148}$$

where the last derivation follows from the definition of $\mathbf{R}(X^*_s, z)$. In the same manner, we get that

$$\mathbb{E}\left[\mathbf{U}_{z,x}(X^*_t) \mid \mathcal{F}\right] \ge \mathbb{E}_s\left[\mathbf{R}(X^*_s, z) \mid \mathcal{F}\right] - \theta \tag{149}$$

for every $z \notin H(x)$.

Since X^*_0 is $(400, R)$-nice at w, and $x \in B_R(w)$, we have that $|H(x)| \le 400\zeta\Delta$. This observation and (148), (149), (142) yield that there exists $C' > 0$ such that

$$\begin{aligned}\mathbf{Pr}\left[\left|\mathbf{R}(X^*_t, x) - \exp\left(-\frac{\lambda}{1+\lambda}\sum_{z \in N(x)} \mathbb{E}_s\left[\mathbf{R}(X^*_s, z) \mid \mathcal{F}\right]\right)\right| \ge 7(\theta + 400\zeta + 3\gamma)\right] \\ \le \exp\left(-C'\Delta\right),\end{aligned} \tag{150}$$

where we use the fact $\frac{\lambda}{1+\lambda}\Delta < e$ and θ, ζ, γ are sufficiently small.

To get from (X_t^*) to (X_t) we use Lemma 29 with parameter γ^3. That is, we have that

$$\mathbf{Pr}\left[\exists s \in I \ |(X_s \oplus X_s^*) \cap S_2(x)| \geq \gamma^3 \Delta\right] \leq \exp\left(-\Delta/C''\right)$$

for some sufficiently large constant $C'' > 0$. This implies that

$$\mathbf{Pr}\left[\exists t \in I \ |\mathbf{R}(X_t^*, x) - \mathbf{R}(X_t, x)| \geq \gamma^2\right] \leq \exp\left(-\Delta/C''\right), \tag{151}$$

since changing any $\Delta\gamma^3$ variables in $S_2(x)$ can only change $\mathbf{R}(X_s^*, x)$ by at most γ^2.

With the same observation we also get that with probability at least $1 - \exp(-\Delta/C'')$ it holds that

$$\left|\int_0^t \mathbf{R}(X_s^*, x) n \exp\left[(s-t)/n\right] ds - \int_0^t \mathbf{R}(X_s, x) n \exp\left[(s-t)/n\right] ds\right| \leq 2\gamma^2. \tag{152}$$

Plugging (151), (152) into (150) and taking appropriate γ, ζ the following is true: There exists $\hat{C} > 0$ such that

$$\mathbf{Pr}\left[\left|\mathbf{R}(X_t, x) - \exp\left(-\frac{\lambda}{1+\lambda} \sum_{r \in N(x)} \mathcal{E}(r)\right)\right| \geq \frac{\eta\epsilon}{20C_0}\right] \leq \exp\left[-\Delta/\hat{C}\right],$$

where

$$\mathcal{E}(r) = \exp[-t/n] \cdot \mathbf{R}(X_0, r) + \int_0^t \mathbf{R}(X_s, r) n \exp\left[(s-t)/n\right] ds.$$

At this point, we remark that the above tail bound holds for a fixed $t \in I$. For our purpose, we need a tail bound which holds for *every* $t \in I$.

Consider a partition of the time interval I into subintervals each of length $\frac{\zeta^3}{200\Delta} n$, where the last part can be of smaller length. Let $T(j)$ be the jth part. Each $z \in S_2(x)$ is updated during the time period $T(j)$ with probability less than $\frac{\zeta^3}{100\Delta}$, independently of the other vertices.

Note that $|S_2(x)| \leq \Delta^2$. Chernoff's bounds imply that with probability at least $1 - \exp(-\Delta\zeta^3)$, the number of vertices in $S_2(x)$ which are updated more than once during the time interval $T(j)$ is at most $\zeta^3\Delta$. Also, changing any $\Delta\zeta^3$ variables in $S_2(x)$ can only change $\mathbf{R}(X_s, x)$ by at most ζ^2.

The lemma follows by taking a union bound over all $T(j)$ for $j \in \{1, \ldots, \lceil 200|I| \Delta/(\zeta^3)\rceil\}$ and all vertices $B_{R/2}(w)$.

11. Conclusions. The work of Weitz [42] was a notable accomplishment in the field of approximate counting/sampling. However a limitation of his approach is that the running time depends exponentially on $\log \Delta$. It is widely believed that the Glauber dynamics has mixing time $O(n \log n)$ for all G of maximum degree Δ when $\lambda < \lambda_c(\Delta)$. However, until now there was little theoretical work to support this conjecture. We give the first such results which analyze the widely used algorithmic approaches of MCMC and loopy BP.

One appealing feature of our work is that it directly ties together with Weitz's approach: Weitz uses decay of correlations on trees to truncate his self-avoiding walk tree, whereas we use decay of correlations to deduce a contracting metric for the path coupling analysis, at least when the chains are at the BP fixed point. We believe this technique of utilizing the principal eigenvector for the BP operator for the path coupling metric will apply to a general class of spin systems, such as 2-spin antiferromagnetic spin systems. (Weitz's algorithm was extended to this class [20].)

We hope that in the future more refined analysis of the local uniformity properties will lead to relaxed girth assumptions. However dealing with very short cycles, such as triangles, will require a new approach since loopy BP no longer seems to be a good estimator of the Gibbs distribution for certain examples.

REFERENCES

[1] R. Bubley and M. Dyer, *Path coupling, Dobrushin uniqueness, and approximate counting*, in Proceedings of the 38th Annual IEEE Symposium on Foundations of Computer Science (FOCS), 1997, pp. 223–231.

[2] V. Chandrasekaran, M. Chertkov, D. Gamarnik, D. Shah, and J. Shin, *Counting independent sets using the Bethe approximation*, SIAM J. Discrete Math., 25 (2011), pp. 1012–1034.

[3] D. Dubhashi and D. Ranjan, *Balls and bins: A study in negative dependence*, Random Structures Algorithms, 13 (1998), pp. 99–124.

[4] M. E. Dyer, A. M. Frieze, T. P. Hayes, and E. Vigoda, *Randomly coloring constant degree graphs*, Random Structures Algorithms, 43 (2013), pp. 181–200.

[5] M. Dyer and C. Greenhill, *On Markov chains for independent sets*, J. Algorithms, 35 (2000), pp. 17–49.

[6] A. Galanis, Q. Ge, D. Štefankovič, E. Vigoda, and L. Yang, *Improved inapproximability results for counting independent sets in the hard-core model*, Random Structures Algorithms, 45 (2014), pp. 78–110.

[7] A. Galanis, D. Štefankovič, and E. Vigoda, *Inapproximability of the partition function for the antiferromagnetic Ising and hard-core models*, Combin. Probab. Comput., 25 (2016), pp. 500–559.

[8] A. Galanis, D. Štefankovič, and E. Vigoda, *Inapproximability for antiferromagnetic spin systems in the tree non-uniqueness region*, J. ACM, 62 (2015), 50.

[9] C. Greenhill, *The complexity of counting colourings and independent sets in sparse graphs and hypergraphs*, Comput. Complexity, 9 (2000), pp. 52–72.

[10] T. P. Hayes, *Local uniformity properties for Glauber dynamics on graph colorings*, Random Structures Algorithms, 43 (2013), pp. 139–180.

[11] T. P. Hayes, *A simple condition implying rapid mixing of single-site dynamics on spin systems*, in Proceedings of the 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS), 2006, pp. 39–46.

[12] T. P. Hayes, J. C. Vera, and E. Vigoda, *Randomly coloring planar graphs with fewer colors than the maximum degree*, Random Structures Algorithms, 47 (2015), pp. 731–759.

[13] T. Hayes and E. Vigoda, *Coupling with the stationary distribution and improved sampling for colorings and independent sets*, Ann. Appl. Probab., 16 (2006), pp. 1297–1318.

[14] T. Heskes, *On the uniqueness of loopy belief propagation fixed points*, Neural Comput., 16 (2004), pp. 2379–2413.

[15] M. Huber, *Approximation algorithms for the normalizing constant of Gibbs distributions*, Ann. Appl. Probab., 25 (2015), pp. 974–985.

[16] A. T. Ihler, J. W. Fisher III, and A. S. Willsky, *Loopy belief propagation: Convergence and effects of message errors*, J. Mach. Learn. Res., 6 (2005), pp. 905–936.

[17] S. Janson, T. Luczak and A. Rucinski, *Random Graphs*, John Wiley and Sons, New York, 2011.

[18] F. P. Kelly, *Stochastic models of computer communication systems*, J. Roy. Statist. Soc. Ser. B, 47 (1985), pp. 379–395.

[19] L. Li, P. Lu, and Y. Yin, *Approximate counting via correlation decay in spin systems*, in Proceedings of the 23rd Annual ACM-SIAM Symposium on Discrete Algorithms (SODA), 2012, pp. 922–940.

[20] L. Li, P. Lu, and Y. Yin, *Correlation decay up to uniqueness in spin systems*, in Proceedings of the 24th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA), 2013, pp. 67–84.

[21] C. Lin, J. Liu, and P. Lu, *A simple FPTAS for counting edge covers*, in Proceedings of the 25th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA), 2014, pp. 341–348.

[22] J. Liu and P. Lu, *FPTAS for #BIS with degree bounds on one side*, in Proceedings of the 47th Annual ACM Symposium on Theory of Computing (STOC), 2015, pp. 549–556.

[23] M. Luby and E. Vigoda, *Fast convergence of the Glauber dynamics for sampling independent sets*, Random Structures Algorithms, 15 (1999), pp. 229–241.

[24] M. Mitzenmacher and E. Upfal, *Probability and Computing: Randomized Algorithms and Probabilistic Analysis*, Cambridge University Press, Cambridge, 2005.

[25] E. Mossel and A. Sly, *Exact thresholds for Ising–Gibbs samplers on general graphs*, Ann. Probab., 41 (2013), pp. 294–328.
[26] E. Mossel, D. Weitz, and N. Wormald, *On the hardness of sampling independent sets beyond the tree threshold*, Probab. Theory Related Fields, 143 (2009), pp. 401–439.
[27] K. P. Murphy, Y. Weiss, and M. I. Jordan, *Loopy belief propagation for approximate inference: An empirical study*, in Proceedings of the 15th Conference on Uncertainty in Artificial Intelligence (UAI), 1999, pp. 467–475.
[28] J. Pearl, *Reverend Bayes on inference engines: A distributed hierarchical approach*, in Proceedings of the 2nd National Conference on Artificial Intelligence (AAAI), 1982, pp. 133–136.
[29] R. Restrepo, J. Shin, P. Tetali, E. Vigoda, and L. Yang, *Improved mixing condition on the grid for counting and sampling independent sets*, Probab. Theory Related Fields, 156 (2013), pp. 75–99.
[30] J. Shin, *The complexity of approximating a Bethe equilibrium*, IEEE Trans. Inform. Theory, 60 (2014), pp. 3959–3969.
[31] A. Sinclair, P. Srivastava, D. Štefankovič, and Y. Yin, *Spatial mixing and the connective constant: Optimal bounds*, Probab. Theory Related Fields, 168 (2017), pp. 153–197.
[32] A. Sinclair, P. Srivastava, and M. Thurley, *Approximation algorithms for two-state anti-ferromagnetic spin systems on bounded degree graphs*, J. Stat. Phys., 155 (2014), pp. 666–686.
[33] A. Sinclair, P. Srivastava, and Y. Yin, *Spatial mixing and approximation algorithms for graphs with bounded connective constant*, in Proceedings of the 54th Annual IEEE Symposium on Foundations of Computer Science (FOCS), 2013, pp. 300–309.
[34] A. Sly, *Computational transition at the uniqueness threshold*, in Proceedings of the 51st Annual IEEE Symposium on Foundations of Computer Science (FOCS), 2010, pp. 287–296.
[35] A. Sly and N. Sun, *The computational hardness of counting in two-spin models on d-regular graphs*, Ann. Probab., 42 (2014), pp. 2383–2416.
[36] D. Štefankovič, S. Vempala, and E. Vigoda, *Adaptive simulated annealing: A near-optimal connection between sampling and counting*, J. ACM, 56 (2009), pp. 1–36.
[37] S. Tatikonda and M. Jordan, *Loopy belief propagation and Gibbs measures*, in Proceedings of the 18th Conference on Uncertainty in Artificial Intelligence (UAI), 2002, pp. 493–500.
[38] L. G. Valiant, *The complexity of enumeration and reliability problems*, SIAM J. Comput., 8 (1979), pp. 410–421.
[39] J. C. Vera, E. Vigoda, and L. Yang, *Improved bounds on the phase transition for the hard-core model in* 2 *dimensions*, SIAM J. Discrete Math., 29 (2015), pp. 1895–1915.
[40] E. Vigoda, *A note on the Glauber dynamics for sampling independent sets*, Electron. J. Combin., 8 (2001), 8.
[41] Y. Weiss, *Correctness of local probability propagation in graphical models with loops*, Neural Comput., 12 (2000), pp. 1–41.
[42] D. Weitz, *Counting independent sets up to the tree threshold*, in Proceedings of the 38th Annual ACM Symposium on Theory of Computing (STOC), 2006, pp. 140–149.